

VANIA GRYZAK GEVERT

**ANÁLISE DE CRÉDITO BANCÁRIO COM O USO DE MODELOS DE
REGRESSÃO LOGÍSTICA, REDES NEURAIS E *SUPPORT VECTOR*
*MACHINE***

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciências, do Programa de Pós-Graduação em Métodos Numéricos em Engenharia, na Área de Concentração em Programação Matemática, Setor de Tecnologia, Departamento de Construção Civil e Setor de Ciências Exatas, Departamento de Matemática da Universidade Federal do Paraná.

Orientador: Prof. Dr. Arinei Carlos Lindbeck da Silva

CURITIBA

2009

TERMO DE APROVAÇÃO

VANIA GRYZAK GEVERT

ANÁLISE DE CRÉDITO BANCÁRIO COM O USO DE MODELOS DE REGRESSÃO LOGÍSTICA, REDES NEURAIS E *SUPPORT VECTOR MACHINE*

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Ciências, do Programa de Pós-Graduação em Métodos Numéricos em Engenharia, na Área de Concentração em Programação Matemática, Setor de Tecnologia, Departamento de Construção Civil e Setor de Ciências Exatas, Departamento de Matemática da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador:

Prof. Dr. Arinei Carlos Lindbeck da Silva
Programa de Pós Graduação em Métodos Numéricos, UFPR

Prof. Dr. Osmar Ambrósio de Souza
Departamento de Matemática, UNICENTRO

Prof. Dr. Volmir Eugênio Wilhem
Programa de Pós Graduação em Métodos Numéricos, UFPR

Curitiba, 26 de agosto de 2009.

Dedico este trabalho ao meu esposo,
pelo seu apoio incondicional mesmo nas
horas mais difíceis, permitindo-me
chegar ao fim de mais esta etapa em
minha vida.

AGRADECIMENTOS

Principalmente agradeço a DEUS, meu grande companheiro, pela força e amparo nesta jornada.

Aos meus pais, Vivaldi e Lindamar, pelo exemplo de vida e por me ensinarem a valorizar a educação.

Ao meu esposo amado Fabiano, que me incentivou na realização desse projeto e que não me deixou fraquejar nos momentos difíceis.

A minha filha Ana Carolina, razão da minha vida, que apesar da pouca idade, soube entender e aceitar todas as vezes que não pudemos estar juntas.

A minha irmã Viviane, minha melhor amiga, que sempre esteve ao meu lado, em todos os momentos de minha vida, compartilhando e torcendo pelo meu sucesso.

A minha segunda família, Almir, Solange e Fabíola, por terem dado o apoio necessário durante todo esse período, compreendendo a importância desta etapa em minha vida.

Ao professor e amigo Arinei Carlos Lindbeck da Silva, pela orientação ao longo desta dissertação, pela enorme paciência, pelo incentivo e confiança e, principalmente, pela amizade.

Aos professores do PPGMNE pelos ensinamentos e apoio durante o curso e em especial ao professor Anselmo Chaves Neto a quem eu aprendi a admirar e a querer bem.

Ao coordenador do curso de Administração das Faculdades Campo Real, Jonatã, por ajustar os horários de aulas conforme as minhas necessidades.

As colegas Luciana, Tânia e Marina, que deixaram de ser simplesmente colegas para serem minhas amigas. Obrigada pelos momentos de descontração e trocas de experiências.

A minha grande amiga e incentivadora Rosana, pelas conversas e desabafos e por estar sempre pronta a me socorrer.

Ao amigo Nico que tanto me incentivou a fazer o Mestrado.

A minha amiga Noidi, por entender a minha ausência e por estar sempre pronta a ajudar.

A Vanessa que disponibilizou o programa de *Support Vector Machines* e dedicou seu tempo em meu auxílio.

Aos colegas de mestrado pelo companheirismo, em especial: Gustavo.

A Maristela pelo seu carinho e paciência.

A todos os colegas do CESEC, pelo companheirismo.

RESUMO

Na área de crédito bancário, ao tomar um empréstimo ou realizar qualquer negócio, o passo seguinte ao preenchimento do cadastro na unidade concedente é a análise de crédito. O uso e a posse de instrumentos que auxiliem na tarefa de classificar possíveis solventes ou insolventes podem tornar-se uma grande vantagem competitiva. Nos dados das grandes empresas, existem muitos conhecimentos úteis para chegar a uma conclusão sobre essa classificação. Para isso, é preciso transformar esses dados em informações valiosas para auxiliar os bancos nessa análise. O presente trabalho foi desenvolvido para prever empresas adimplentes e inadimplentes. Para isso, são analisados o registro de 199 clientes (pessoas jurídicas) de uma agência bancária na cidade de Wenceslau Braz, no Estado do Paraná, utilizando três técnicas: Regressão Logística, Redes Neurais e *Support Vector Machine*. O objetivo é a construção de cada modelo e a comparação do desempenho das três técnicas escolhidas. Os resultados encontrados demonstram uma superioridade de *Support Vector Machine* em relação às outras duas técnicas. Alguns dos motivos para esse sucesso estão relacionados ao fato dessa técnica exibir bom desempenho de generalização em muitas bases de dados reais, ser bem fundamentada matematicamente, a eliminação da possibilidade de mínimos locais pelo processo de treinamento e a existência de poucos parâmetros livres para ajuste.

ABSTRACT

In the area from bank credit to make a loan or to make any credit business, the next step to fill out the register in the allowed unit is to analyse the credit. The use and the ownership from the instruments that help in the task to classify the possible payers and the people that do not pay can become a big competitive advantage. In the data from the big companies there are a lot of useful knowledges to arrive in a conclusion about this classification. For this is necessary to transform these data in valorous informations to help the banks in this analyse. The present work was developed to show the companies that pay and the companies that do not pay their bills. To this are analysed the register from 199 clients (partnership) from a bank agency in the Wenceslau Braz city, in the Parana State, using three techniques: Logistics Regression, Neural Networks and Support Vector Machine. The object is a building from each model and the comparative three techniques chosen. The found results show a superiority of Support Vector Machine in relation to the other to techniques. Some of the reasons to this success are related to the fact of this technique show a good performance of generalization in a lot of reals data bases, to be well fundamental mathematically, the elimination of the possibility from the local minimum for the training process and the existence of a few free parameters to adjust.

LISTA DE TABELAS

TABELA 4.1	PERFORMANCE DAS FUNÇÕES <i>KERNEL</i> E SEUS PARÂMETROS	94
TABELA 4.2	FUNÇÕES <i>KERNEL</i> TREINADAS E SEUS PARÂMETROS	96
TABELA 4.3	PERFORMANCE DAS FUNÇÕES <i>KERNEL</i> , DESVIO PADRÃO E TEMPO COMPUTACIONAL	96
TABELA 4.4	FUNÇÕES <i>KERNEL</i> TREINADAS E SEUS PARÂMETROS - AF	108
TABELA 4.5	PERFORMANCE DAS FUNÇÕES <i>KERNEL</i> , DESVIO PADRÃO E TEMPO COMPUTACIONAL - AF.....	108

LISTA DE FIGURAS

FIGURA 2.1	RELAÇÃO ENTRE O BANCO E O TOMADOR	21
FIGURA 3.1	HIERARQUIA DO APRENDIZADO.....	44
FIGURA 3.2	CONSTITUINTES DA CÉLULA NEURAL	45
FIGURA 3.3	CONEXÕES EM UMA REDE NEURAL ARTIFICIAL	48
FIGURA 3.4	POSSIBILIDADES DE HIPERPLANOS SEPARADORES	51
FIGURA 3.5	FORMAÇÃO DO HIPERPLANO DE SEPARAÇÃO ATRAVÉS DOS VETORES DE SUPORTE.....	52
FIGURA 3.6	INTERPRETAÇÃO DAS DISTÂNCIAS ALGÉBRICAS DE PONTOS ATÉ O HIPERPLANO ÓTIMO PARA UM CASO BIDIMENSIONAL	53
FIGURA 3.7(a)	PONTO DADO DENTRO DA REGIÃO DE SEPARAÇÃO, MAS DO LADO CORRETO DA SUPERFÍCIE DE DECISÃO	60
FIGURA 3.7(b)	PONTO DADO NO LADO ERRADO DA SUPERFÍCIE DE DECISÃO...	60
FIGURA 3.8	MAPEAMENTO DOS DADOS DO ESPAÇO DE ENTRADA NO ESPAÇO <i>FEATURE</i>	64
FIGURA 3.9	MAPEAMENTO REALIZADO PELA FUNÇÃO <i>KERNEL</i> EM UM ESPAÇO DE DIMENSÃO MAIOR	65
FIGURA 3.10	SEPARAÇÃO DOS DADOS.....	69
FIGURA 3.11	SEPARAÇÃO DOS DADOS EM SUB-REGIÕES.....	70
FIGURA 3.12	FLUXOGRAMA DA ESCOLHA DO PRIMEIRO PONTO	77
FIGURA 3.13	FLUXOGRAMA DA ESCOLHA DO SEGUNDO PONTO	78
FIGURA 3.14	FLUXOGRAMA DA OTIMIZAÇÃO DOS PONTOS ESCOLHIDOS	79
FIGURA 4.1	MATRIZ DE CONFUSÃO GERADA PELO <i>SOFTWARE STATISTICA</i>	87
FIGURA 4.2	PÁGINA INICIAL DO PROGRAMA	92
FIGURA 4.3	TELA MOSTRANDO O NÚMERO DE PONTOS E A DIMENSÃO	93
FIGURA 4.4	TELA DE ESCOLHA DA FUNÇÃO <i>KERNEL</i> E DOS PARÂMETROS ...	93
FIGURA 4.5	PARTE DO ARQUIVO RESULTANTE APÓS O SVM	95
FIGURA 4.6	MATRIZ DE CONFUSÃO GERADA PELO <i>SOFTWARE</i> <i>STATISTICA</i> – AF	103

LISTA DE GRÁFICOS

GRÁFICO 2.1	REPRESENTAÇÃO DAS IDADES DAS EMPRESAS INSOLVENTES...	25
GRÁFICO 3.1	FUNÇÃO LOGÍSTICA OU SIGMÓIDE	40

LISTA DE QUADROS

QUADRO 2.1	PRECISÃO DA CLASSIFICAÇÃO DOS MODELOS	32
QUADRO 2.2	PORCENTAGEM DE PRECISÃO NA CLASSIFICAÇÃO DOS MODELOS CONSTRUÍDOS (LITERATURA PESQUISADA)	33
QUADRO 3.1	EXEMPLO DE CLASSIFICAÇÃO DOS 200 CASOS DE PORTADORES E NÃO PORTADORES DE CELULAR	42
QUADRO 3.2	HISTÓRICO DAS REDES NEURAIS ARTIFICIAIS EM ORDEM CRONOLÓGICA	46
QUADRO 4.1	RESPOSTAS DO QUESTIONÁRIO APLICADO PELO BANCO ÀS EMPRESAS	84
QUADRO 4.2	BETAS ESTIMADOS PELO MÉTODO DA MÁXIMA VEROSSIMILHANÇA	86
QUADRO 4.3	RESULTADOS PARCIAIS DA TÉCNICA DE REGRESSÃO LOGÍSTICA.....	88
QUADRO 4.4	RESULTADO FINAL DA REGRESSÃO LOGÍSTICA.....	88
QUADRO 4.5	PERFORMANCE DAS REDES TREINADAS	90
QUADRO 4.6	PESOS E BIAS DA REDE Nº 5	91
QUADRO 4.7	COMPARAÇÃO ENTRE AS TÉCNICAS	98
QUADRO 4.8	VARIÁVEIS RESULTANTES	100
QUADRO 4.9	AUTOVALORES E % DE VARIÂNCIA EXPLICADA.....	100
QUADRO 4.10	PESOS FATORIAIS ROTACIONADOS	101
QUADRO 4.11	ESCORES FATORIAIS ROTACIONADOS	102
QUADRO 4.12	BETAS ESTIMADOS PELO MÉTODO DA MÁXIMA VEROSSIMILHANÇA – AF.....	103
QUADRO 4.13	RESULTADOS PARCIAIS DA TÉCNICA DE REGRESSÃO LOGÍSTICA – AF	105
QUADRO 4.14	RESULTADO FINAL DA REGRESSÃO LOGÍSTICA - AF.....	105
QUADRO 4.15	PERFORMANCE DAS REDES TREINADAS – AF	106
QUADRO 4.16	PESOS E BIAS DA REDE Nº 6 – AF	107
QUADRO 4.17	COMPARAÇÃO ENTRE AS TÉCNICAS – AF	110
QUADRO 4.18	COMPARAÇÃO ENTRE AS TÉCNICAS ANTES E DEPOIS DA AF..	110

SUMÁRIO

LISTA DE TABELAS	8
LISTA DE FIGURAS	9
LISTA DE GRÁFICOS	10
LISTA DE QUADROS	11
 CAPÍTULO I - INTRODUÇÃO	 14
1.1 TEMA E OBJETIVOS	15
1.1.1 Objetivo Geral	16
1.1.2 Objetivos Específicos	16
1.2 JUSTIFICATIVA	17
1.3 ESTRUTURA DO TRABALHO	18
 CAPÍTULO II – REFERENCIAL TEÓRICO	 19
2.1 INTRODUÇÃO	19
2.2 BANCOS	20
2.3 CRÉDITO	20
2.4 RISCO DE CRÉDITO	23
2.5 OS CINCO C'S DO CRÉDITO	24
2.5.1 Caráter	24
2.5.2 Capacidade	25
2.5.3 Capital	26
2.5.4 Condições	26
2.5.5 Colateral	27
2.6 MODELOS DE PREVISÃO DE RISCO DE CRÉDITO	28
2.7 REVISÃO BIBLIOGRÁFICA DAS APLICAÇÕES DOS MODELOS DE PREVISÃO	30
 CAPÍTULO III - TÉCNICAS UTILIZADAS	 35
3.1 INTRODUÇÃO	35
3.2 REGRESSÃO LOGÍSTICA	36
3.2.1 Modelos e Técnicas Estatísticas	36
3.2.2 Análise Discriminante x Regressão Logística	38
3.2.3 Regressão Logística	39
3.2.4 O Modelo Matemático Logístico	40
3.2.5 Transformação Logit	41
3.2.6 Ajuste do Modelo de Regressão Logística	41
3.2.7 Verificação do Ajuste	42
3.3 REDES NEURAIS	43
3.3.1 Aprendizagem de máquina	43
3.3.2 Uma introdução às Redes Neurais	44
3.3.3 Histórico das Redes Neurais Artificiais	44
3.3.4 Conceitos, características e estrutura da RNA	46
3.4 <i>SUPPORT VECTOR MACHINES (SVM)</i>	50
3.4.1 Hiperplano ótimo para padrões linearmente separáveis	50
3.4.2 Otimização Quadrática para Encontrar o Hiperplano Ótimo	55
3.4.3 Hiperplano ótimo para padrões não-separáveis	59
3.4.4 Funções <i>Kernel</i>	63
3.4.5 Implementação ou Soluções do Modelo de Programação Quadrática do SVM	70
3.4.5.1 SMO (<i>Sequential Minimal Optimization</i>)	71

CAPÍTULO IV - METODOLOGIA E RESULTADOS OBTIDOS.....	80
4.1 DELIMITAÇÃO DO PROBLEMA	80
4.2 SELEÇÃO DA AMOSTRA.....	81
4.3 PADRONIZAÇÃO OU NORMALIZAÇÃO DOS DADOS.....	84
4.4 TREINAMENTO E TESTE.....	85
4.5 TÉCNICA DE REGRESSÃO LOGÍSTICA	86
4.5.1 Estimativa dos Parâmetros.....	86
4.5.2 Classificações Corretas.....	87
4.5.3 Equações da Logit Estimada e do Modelo Estimado	87
4.5.4 Resultados da Regressão Logística	88
4.6 TÉCNICA REDES NEURAIIS	89
4.6.1 Modelo Implementado.....	89
4.6.2 Resultados Obtidos Redes Neurais	90
4.6.3 Pesos e <i>bias</i>	91
4.7 TÉCNICA <i>SUPPORT VECTOR MACHINES (SVM's)</i>	92
4.7.1 Testando diferentes <i>kernels</i>	94
4.7.2 Resultados obtidos	95
4.8 COMPARAÇÃO ENTRE AS TÉCNICAS	97
4.9 TREINAMENTO COM ANÁLISE FATORIAL.....	98
4.9.1 Redução de dimensionalidade.....	98
4.9.2 Autovalores e variância explicada	99
4.9.3 Pesos Fatoriais e Escores Fatoriais	101
4.9.4 Matriz de treinamento e teste.....	102
4.10 TÉCNICAS UTILIZADAS	102
4.10.1 Regressão Logística.....	103
4.10.2 Estimativa dos Parâmetros.....	103
4.10.3 Classificações Corretas.....	103
4.10.4 Equações da Logit Estimada e do Modelo Estimado	104
4.10.5 Resultados da Regressão Logística	104
4.10.6 Resultado final da Regressão Logística - AF.....	105
4.11 TÉCNICA REDES NEURAIIS	106
4.11.1 Resultados Obtidos Redes Neurais	106
4.11.2 Pesos e <i>bias</i>	107
4.12 TÉCNICA <i>SUPPORT VECTOR MACHINES (SVM's)</i>	107
4.13 COMPARAÇÃO ENTRE AS TÉCNICAS	109
4.14 RESULTADO FINAL/COMPARAÇÃO ENTRE AS TÉCNICAS	110
CAPÍTULO V - CONCLUSÕES E SUGESTÕES DE TRABALHOS FUTUROS	111
5.1 CONCLUSÕES	111
5.2 TRABALHOS FUTUROS	113
REFERÊNCIAS	114
APÊNDICE A – QUESTIONÁRIO	120
APÊNDICE B – ANÁLISE FATORIAL (AF) E ANÁLISE DE COMPONENTES PRINCIPAIS (ACP)	121
APÊNDICE C – AUTOVALORES E VARIÂNCIA EXPLICADA.....	125
APÊNDICE D – GRÁFICO DOS 20 AUTOVALORES	126
APÊNDICE E – COMUNALIDADES.....	127
APÊNDICE F – MATRIZ DE CORRELAÇÃO RESIDUAL.....	128
APÊNDICE G – ESCORES FATORIAIS.....	129

CAPÍTULO I

1 INTRODUÇÃO

Os primeiros anos do século XXI estão mostrando a dimensão da evolução tecnológica que o homem alcançou e ainda almeja alcançar. A quantidade de informações e dados que se recebe por dia através da televisão, dos *e-mails*, dos *sites* da Internet, telefones celulares e rádios é algo nunca visto. É essa explosão de dados que permite à tecnologia crescer de forma cada vez mais rápida (CUNICO, 2005).

Com recursos da informática, as empresas podem acumular uma quantidade de dados das mais variadas fontes: informações comerciais; industriais; relatórios de vendas; de custos; de demanda, enfim, em quase todas as áreas de aplicação, como Economia, Biologia, Medicina, entre outras.

Por outro lado, somente possuir dados não basta. Então, a partir da década de 80, algumas empresas começaram a descobrir que, em meio a essa imensa quantidade de dados, podiam se esconder informações valiosas (CUNICO, 2005). Percebe-se a necessidade de transformar os dados armazenados em informação.

Para a construção do conhecimento científico acerca dos dados armazenados, necessita-se de ferramentas que auxiliem a leitura, seleção e a forma com que esses dados serão processados e convertidos em informações.

Um dado se transforma em informação quando ganha um significado para seu utilizador; caso contrário, continua sendo um simples dado.

Então surgiu a pergunta: “O que fazer com esses dados para obter determinadas informações?”

Métodos estatísticos têm sido exaustivamente utilizados para dar sentido a estes dados. Porém, novas metodologias têm sido estudadas, dentre elas, *Data Mining*, Redes Neurais, Algoritmos Genéticos, *Support Vector Machine*, entre outras.

As instituições financeiras possuem dados a respeito das empresas que chegam até elas para obter um empréstimo ou financiamento. É necessário no momento da tomada de decisão da concessão de crédito realizar uma análise desta empresa, ou seja, a análise de crédito.

Análise de crédito é um estudo da situação financeira do indivíduo no contrato bancário e de sua capacidade de endividamento e liquidez; ou seja, verifica se a pessoa tem condições de pagar a dívida que está a contrair, ou no mínimo garantir com outro bem o custo nas condições contratadas.

Nesta análise, ainda são levadas em conta informações obtidas através de várias fontes externas como Serasa, SPC (Serviço de Proteção ao Crédito) que representam dados importantes para análise.

A intenção de qualquer instituição de crédito é que os valores emprestados retornem no prazo combinado, possibilitando a realização de novos negócios com rendimento melhor do que pendências nos tribunais de cobrança de devedores inadimplentes, que podem levar muito tempo sem solução.

Os bancos precisam distinguir, entre seus clientes, quais pagarão a dívida contraída, e quais não honrarão o compromisso assumido.

Segundo Guimarães e Chaves Neto (2002), quando a empresa tem à sua disposição uma regra de reconhecimento de padrões e classificação que indique previamente a chance de inadimplência de um futuro cliente, a decisão de concessão de crédito fica muito facilitada, podendo-se utilizar argumentos quantitativos em substituição aos argumentos subjetivos e decidir com maior confiança.

Em muitas empresas, a avaliação de crédito é feita com base em uma enorme variedade de informações vindas das mais variadas fontes. Os gerentes analisam estas informações de forma subjetiva e muitas vezes não conseguem descrever o processo de tomada de decisões, embora consigam apontar os fatores que influenciam nestas decisões. Além disso, estes ambientes são dinâmicos, com constantes alterações, onde as decisões devem ser tomadas rapidamente (MENDES FILHO *et al.*, 1996).

Surge assim a necessidade de se ter ferramentas que auxiliem o analista financeiro a tomar decisões na concessão ou não do crédito.

1.1 TEMA E OBJETIVOS

Este trabalho aborda três temas principais: crédito, risco de crédito e modelos de previsão.

Para isso, é preciso conhecer os processos de Análise de Crédito e os métodos já desenvolvidos para tal finalidade. Dentre as técnicas utilizadas, pode-se citar Árvores de

Decisão, Análise de Discriminante, Redes Neurais, Algoritmo Genético, Regressão Logística, Indução de Regras e *Support Vector Machine*, optando-se pela utilização de três delas: Regressão Logística, Redes Neurais e *Support Vector Machine*. Estas técnicas permitem o reconhecimento de padrões e preveem futuros diagnósticos.

1.1.1 Objetivo geral

Comparar o uso das técnicas Regressão Logística, Redes Neurais e *Support Vector Machine* para análise da concessão de crédito à pessoas jurídicas.

1.1.2 Objetivos específicos

Além do objetivo geral, são estabelecidos os seguintes objetivos específicos:

- Conhecer alguns processos de Análise de Crédito e os modelos de classificação já desenvolvidos;
- Desenvolver três modelos de previsão de risco de crédito, também denominados *credit scoring*, mediante o uso das técnicas Regressão Logística, Redes Neurais e *Support Vector Machine* antes e após a Análise Fatorial;
- Classificar novas empresas através do modelo obtido como prováveis adimplentes ou inadimplentes;
- Comparar desempenhos, vantagens e desvantagens dos modelos construídos e identificar qual modelo apresentou-se como o mais indicado para a classificação das empresas;
- Propor um modelo para a classificação das empresas;
- Realizar uma Análise Fatorial para reduzir a dimensionalidade do problema de análise de crédito bancário;
- Determinar as taxas de acerto dos modelos construídos após a AF;
- Comparar as técnicas utilizadas antes e após a Análise Fatorial;
- Analisar se as metodologias poderão ajudar, e se elas beneficiarão o trabalho de concessão de créditos, facilitando o gerenciamento e reduzindo a inadimplência através da

identificação e classificação dos prováveis “bons” e “maus” pagadores, aumentando com isto a lucratividade e evitando perdas financeiras acarretadas pela inadimplência.

1.2 JUSTIFICATIVA

Ao tomar um empréstimo ou realizar qualquer negócio a crédito, seja pessoal, imobiliário, agrícola, ou até mesmo um cartão de crédito, o passo seguinte ao preenchimento do cadastro na instituição concedente é a análise de crédito.

Observando que as instituições financeiras realizam análise de crédito para fornecer um empréstimo ou realizar qualquer negócio, sentiu-se a necessidade de auxiliar no trabalho dessa concessão bancária.

Segundo Securato (2002), na área financeira das indústrias, do comércio e de demais atividades econômicas, os homens de finanças estão constantemente envolvidos com tomadas de decisão que podem representar o sucesso ou fracasso próprio, de um projeto específico ou mesmo da empresa. Conceder ou não crédito, tomar recursos ou aplicá-los são decisões constantes a que estão sujeitos os executivos financeiros, qualquer que seja o negócio da empresa.

Para Silva (2003), as técnicas estatísticas multivariadas têm sido consideradas como ferramentas bastante poderosas na administração do risco de inadimplência existente na concessão de crédito. Para a construção de um modelo de previsão de risco de crédito é importante identificar qual a técnica estatística mais eficiente para modelar os dados da empresa de forma a conseguir a melhor previsão de comportamento dos clientes.

Nenhum modelo consegue precisão absoluta, ou seja, acertar totalmente as previsões. Sabendo disto, qualquer avanço em acuracidade da previsão gera ganhos financeiros para a instituição concedente. Daí vem o interesse de analisar diferentes tipos de modelos e apontar quais apresentam uma maior precisão.

Na literatura pesquisada, principalmente no Brasil, encontram-se poucos estudos que abordam *Support Vector Machine* como ferramenta de modelos de *credit scoring*. Em contrapartida, Redes Neurais e Regressão Logística são largamente empregadas neste tipo de problema. Por esta razão, julgou-se oportuno apresentar as três técnicas para a utilização em um mesmo banco de dados e comparar suas vantagens e desvantagens.

Os dados utilizados neste trabalho foram obtidos de uma agência bancária na cidade de Wenceslau Braz, interior do Paraná, que trabalha na concessão de crédito de pessoas

jurídicas (empresas). O público alvo são pequenas e médias empresas, e a instituição financeira coloca à disposição linhas de crédito tanto para capital de giro quanto para investimentos.

1.3 ESTRUTURA DO TRABALHO

Este trabalho foi dividido em cinco capítulos, e seus conteúdos estão descritos na sequência:

- Capítulo 1: Introdução ao assunto, justificativa do problema, tema e objetivos a serem alcançados.
- Capítulo 2: Conceitos fundamentais de crédito, risco de crédito, modelos de risco de crédito e revisão bibliográfica das aplicações dos modelos de previsão.
- Capítulo 3: Descrição detalhada da fundamentação teórica e matemática das técnicas envolvidas na solução do problema.
- Capítulo 4: Descrição da aplicação das técnicas, a metodologia, os resultados obtidos e a comparação entre as técnicas.
- Capítulo 5: Apresentação das conclusões realizadas a partir do capítulo anterior, além de sugestões para a realização de trabalhos futuros.

CAPÍTULO II

2 REFERENCIAL TEÓRICO

2.1 INTRODUÇÃO

Este capítulo apresenta uma visão geral sobre os aspectos conceituais do crédito, que constituirão o alicerce básico para compreensão das funções de uma área de crédito, para a avaliação do risco das empresas, para estruturação de empréstimos e financiamentos e para a gestão do crédito. A Análise de Crédito consiste em interpretar e analisar de forma conjunta os dados disponíveis de uma empresa e emitir parecer sobre sua situação econômico-financeira. Este trabalho focará especificamente em análise de crédito bancário à pessoa jurídica (pequenas e médias empresas).

As empresas precisam de recursos para atender às suas necessidades de investimento e de capital de giro, e às instituições financeiras cumpre a função de supridores desses recursos, no desempenho de seus papéis de intermediação financeira (SILVA, 2003).

O crédito é um instrumento tão importante na vida das pessoas e das empresas que já é habitual as lojas estabelecerem preços para venda de seus produtos para pagamentos em parcelas. Para as instituições de crédito em geral, cujo principal produto é o dinheiro, não há a possibilidade de venda à vista, o que torna o crédito o próprio negócio da instituição. Empresas que necessitam comprar equipamentos ou matéria-prima podem valer-se do crédito como uma forma de acionar o progresso de suas atividades, gerando impostos e empregos no país, entre outros benefícios.

O objetivo final da instituição ou banco é que seja efetuado o pagamento do empréstimo. Porém, nem sempre isso é possível, levando os tomadores de empréstimos à inadimplência. Alguns métodos quantitativos foram desenvolvidos por estudiosos na área para prever e classificar prováveis adimplentes e inadimplentes. Como o trabalho a ser realizado pretende classificar empresas nestas duas classes, é necessário conhecimento sobre os diversos modelos desenvolvidos, bem como suas vantagens e limitações.

2.2 BANCOS

O Sistema Financeiro Nacional pode ser dividido em dois subsistemas, ou seja, um normativo e outro operacional. O subsistema normativo é responsável pela normatização e controle das instituições que operam no mercado. É composto, basicamente, pelo Conselho Monetário Nacional, Banco Central do Brasil, Banco do Brasil, Banco Nacional de Desenvolvimento Econômico e Social e Comissão de Valores Mobiliários. O subsistema operativo é composto pelas instituições financeiras (bancárias e não bancárias) que operam diretamente na intermediação financeira (SILVA, 2003).

Um banco é, essencialmente, um intermediário financeiro que exerce o papel de aproximar os agentes econômicos (famílias, empresas e governo) com situação orçamentária superavitária dos agentes econômicos com situação deficitária, visando atender às necessidades das duas partes.

As funções de captar e aplicar recursos, associadas à prestação de serviços, é que propiciam o retorno para os acionistas dos bancos. Associado ao mesmo retorno, está o risco a que esses acionistas estão expostos.

Muitas pessoas podem pensar que os bancos emprestam apenas seus recursos para atender as necessidades de consumo ou de investimento de seus clientes. Na verdade, um banco utiliza recursos próprios e recursos de terceiros. Entre os recursos de terceiros, estão os depósitos efetuados pelos seus clientes.

Os depósitos cumprem importante papel na intermediação financeira e se constituem numa grande fonte de fundos para os bancos atenderem às necessidades de seus clientes tomadores de recursos.

A função de crédito consiste em avaliar a capacidade de pagamento do tomador, visando assegurar a reputação e a solidez do banco. Dar segurança aos depositantes sobre os valores confiados e garantir o retorno aos acionistas são objetivos fundamentais.

A intermediação financeira é a principal atividade de um banco. Emprestar dinheiro e financiar a aquisição de bens é sua principal fonte de renda (SILVA, 2003).

2.3 CRÉDITO

Crédito é uma palavra que deriva de *credare*: expressão latina que significa *confiar* ou *acreditar*.

Segundo Berni (1999), o crédito significa confiança de que algum evento esperado venha a acontecer.

Para Schrickel (2000), crédito é todo ato de vontade ou disposição de alguém de destacar ou ceder, temporariamente, parte de seu patrimônio a um terceiro, com a expectativa de que esta parcela volte a sua posse integralmente, após decorrido um determinado tempo previamente estipulado.

Mas, dependendo do contexto do qual esteja se tratando, tem vários significados. Num sentido restrito e específico, crédito consiste na entrega de um valor presente mediante uma promessa de pagamento¹. Numa loja de informática, uma venda a crédito é caracterizada pela entrega de mercadoria (computador, *notebook*, artigos de informática, entre outros) ao cliente, mediante uma promessa de pagamento, em uma ou mais parcelas, num prazo futuro, definido de comum acordo entre as partes. Em um banco, que tem a intermediação financeira como sua principal atividade, o crédito consiste em colocar à disposição do cliente (tomador de recursos) certo valor sob a forma de empréstimo ou financiamento, mediante uma promessa de pagamento numa data futura. Na verdade, o banco está comprando uma promessa de pagamento, pagando ao tomador (vendedor) um determinado valor para, no futuro, receber um valor maior. A Figura 2.1 ilustra a relação entre o banco e o tomador.

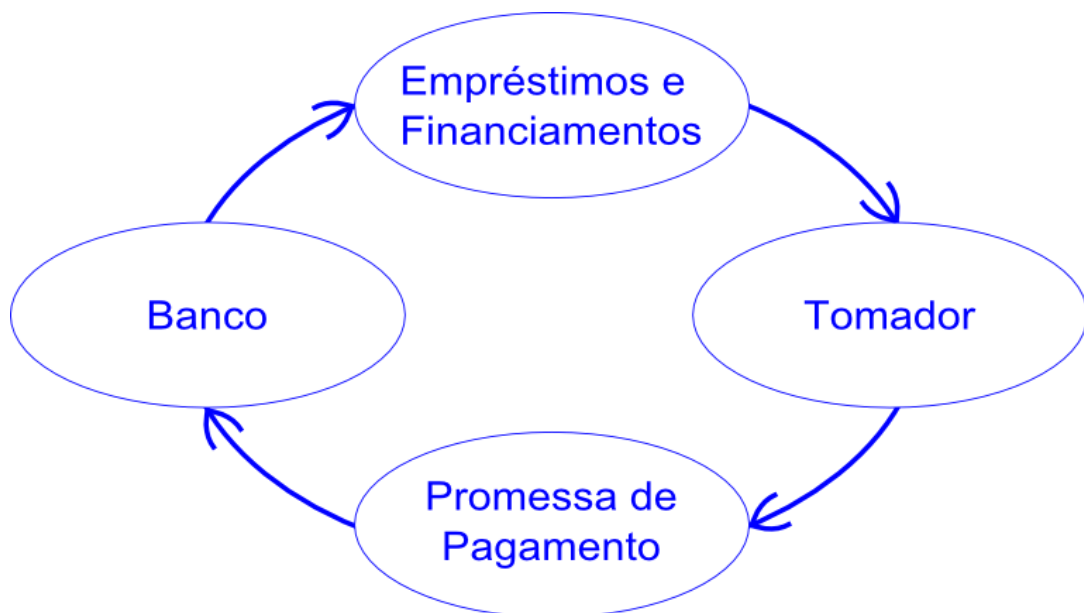


FIGURA 2.1: RELAÇÃO ENTRE O BANCO E O TOMADOR
FONTE: A autora (2009)

¹ Encontramos, com frequência, a definição de crédito como sendo algo do tipo “... é a confiança de que a promessa de pagamento será honrada”. Entendemos que a confiança é um elemento necessário, porém não é suficiente para uma decisão de crédito (SILVA, 2003).

No comércio, de um modo geral, o crédito assume o papel de facilitador da venda. Possibilita ao cliente adquirir o bem para atender sua necessidade, ao mesmo tempo em que incrementa às vendas do comerciante. Em algumas atividades comerciais, o ganho no financiamento das mercadorias chega a ser maior que a própria margem praticada em sua atividade principal.

Na indústria, o crédito também assume o papel de facilitador de venda. Um fabricante de produtos hospitalares, por exemplo, pode abrir linhas de crédito para vendas de seus produtos e com isso possibilitar a vários médicos, clínicas e hospitais a aquisição de seus equipamentos. Caso não houvesse a alternativa de crédito, a quantidade de compradores poderia ser muito menor e, conseqüentemente, o lucro do fabricante também seria reduzido.

Em finanças o vocábulo crédito define um instrumento de política de negócios a ser utilizado por uma empresa comercial ou industrial na venda a prazo de seus produtos, ou por um banco comercial, por exemplo, na concessão de empréstimo, financiamento ou fiança.

Segundo Silva (2003), o crédito cumpre um importante papel econômico e social, a saber:

- Possibilita às empresas aumentar seu nível de atividade;
- Estimula o consumo estimulando a demanda;
- Ajuda as pessoas a obter moradia, bens e até alimento; e,
- Facilita a execução de projetos para os quais as empresas não disponham de recursos próprios suficientes.

Negar créditos é sempre mais cômodo que aprová-los; mas o banco não vive de créditos negados (SCHRICKEL,1998).

O crédito que alguém dispõe é sua capacidade de obter dinheiro, mercadoria ou serviço mediante compromisso de pagamento num prazo tratado.

Toda a atividade do sistema financeiro está baseada em expectativas, que se realizarão, ou não, no decorrer do tempo. As incertezas quanto ao futuro traduzem-se em riscos para o tomador e prestador. Quanto maior for a distância no tempo entre a concessão do crédito e o pagamento do valor emprestado, maior será o risco. É importante frisar que crédito envolve valores e prazos, e estará associado a um nível de risco. O que se pode fazer é procurar minimizá-lo pela análise segura do negócio e pelo conhecimento que deve-se ter a respeito das pessoas com quem se negocia.

2.4 RISCO DE CRÉDITO

Sumariamente, risco significa incerteza, imponderável, imprevisível, e estes, a incerteza, a imponderabilidade e a imprevisibilidade situam-se, necessária e unicamente, no futuro. Assim, embora a análise de crédito deva lidar com eventos passados do tomador de empréstimos (a análise histórica), as decisões de crédito devem considerar primordialmente o futuro desse mesmo tomador.

A imprevisibilidade quanto ao futuro - o risco - deve ser adequada e suficientemente ponderada, a fim de minimizar os riscos expostos da instituição financeira. Afinal, uma instituição financeira, ou melhor, qualquer instituição financeira, excluídas outras considerações de natureza político-filosóficas, é uma empresa como tantas outras, e tem certos objetivos a perseguir e alcançar que, de resto, se resumem em sua própria manutenção e sobrevivência e interação no meio em que se predispõe a interagir.

Conforme Paiva (1997), para confiar-se em pessoas, precisa-se conhecê-las. E, para conhecê-las, precisa-se basicamente de tempo e informação. Salienta, também, que a confiança não é absoluta. Confia-se diferentemente em pessoas diferentes. Da mesma forma os bancos devem ter informações suficientes para acreditar ser o cliente merecedor de algum crédito.

Existem muitas formas de se conhecer alguém. Uma delas, e muito importante, é o convívio constante e duradouro. Contudo, o clima de confiança, para se instalar em casos como esse, demanda um tempo que não é disposto, se a intenção é ser ágil em um mercado altamente competitivo.

Por isso, é necessário o desenvolvimento de técnicas que antecipem o conhecimento que se deve ter a respeito dos clientes. Ao conjunto dessas técnicas, que se somam e se complementam, dá-se o nome de Análise de Crédito.

Essa forma de análise torna a decisão sobre “confiar ou não no cliente” menos empírica e mais técnica. Assim, ainda que o elemento humano continue desempenhando importante papel, a sensibilidade do analista entra no processo como um complemento, uma variável ponderativa que, aliada à análise técnica, embasa a final declaração do conceito e a decisão sobre conceder ou não crédito ao cliente.

Blatt (1999) define Risco de Crédito como a possibilidade de que aquele que concedeu o crédito não o receba do devedor na época e/ou nas condições combinadas.

Na sequência de análise que se usa para determinar o risco de crédito, toma-se como base as metodologias tradicionais, e, normalmente, levam-se em conta os cinco aspectos do crédito, conhecidos como os cinco C's do crédito.

2.5 OS CINCO C'S DO CRÉDITO

Entre vários autores da literatura que se dedicaram ao assunto, não há consenso sobre quantos cês compõem o crédito. Uns listam quatro, outros entendem que são cinco. Outros, ainda, dizem oito os cês do crédito. Todos, entretanto, são unânimes quanto aos quatro primeiros (caráter, capacidade, condições e capital).

Pode-se encontrar na literatura que trata de crédito, normalmente em estudos de administração de contas a receber, os cinco C's do Crédito enumerados pelos autores Weston e Brigham². São eles: caráter, capacidade, capital, condições e colateral. A análise destas características embasa metodologias utilizadas no processo de avaliação de crédito em todo o mundo.

2.5.1 Caráter

O caráter refere-se à intenção de pagar. Está ligado à honestidade, boa-fé, idoneidade. Diz respeito à intenção do cliente em cumprir a promessa de pagar o valor que lhe foi confiado. Em outras palavras, é a disposição do indivíduo de pagar o empréstimo na data do vencimento e nas condições acordadas. E essa disposição varia de pessoa a pessoa. Um devedor pode chegar a se desfazer de bens essenciais para pagar seus compromissos. Outro, porém, pode não querer fazer qualquer esforço para tal. Analisar o caráter do cliente deve ser o primeiro passo. E como fazer isso? (SILVA, 2003)

Deve-se levantar o histórico de pontualidade do cliente, recorrendo aos registros internos do Banco, buscando informações junto às entidades de proteção ao crédito, como Serasa e SPC (Serviço de Proteção ao Crédito), entre outras, nos cartórios de protestos e em outras fontes disponíveis.

² WESTON, J. Fred, BRIGHAM, Eugene P. *Managerial Finance*. New York: Holt, 1972.

O prestador deve construir um conjunto de informações sobre a qualificação e identificação do indivíduo ou da empresa e sobre as experiências de outros credores em suas relações comerciais com o mesmo. Cabe enfatizar, entretanto, que um indivíduo ou uma empresa pode atrasar um pagamento, ou deixar de pagar, em razão de não dispor de recursos, o que não é decorrência necessariamente de seu caráter. Isto porque é mais importante conhecer quem ele é do que o que ele faz. Dessa forma, a identificação do conjunto de boas ou más qualidades de um indivíduo ou empresa, em face do hábito de pagar suas contas, é tarefa difícil.

2.5.2 Capacidade

A capacidade refere-se à habilidade, à competência empresarial do indivíduo ou do grupo de indivíduos, e ao potencial de produção, administração e comercialização da empresa (SILVA, 2003).

Alguns estudos indicam que empresas tradicionais têm quadros gestores mais experientes no mercado e no setor que atuam, ou seja, a idade pode, certamente, indicar um maior nível de experiência. Usualmente, o sucesso do negócio também está relacionado com o nível de escolaridade do empresário e dos demais membros da empresa.

A análise da idade das empresas propicia-nos importantes informações, de forma que, num conjunto de grande número de empresas insolventes, cuja amostra está representada no Gráfico 2.1, constatamos que a grande maioria das insolvências ocorreu na faixa de três a seis anos da data da fundação, enquanto mais de 50% delas se concentraram na faixa de até nove anos de idade.

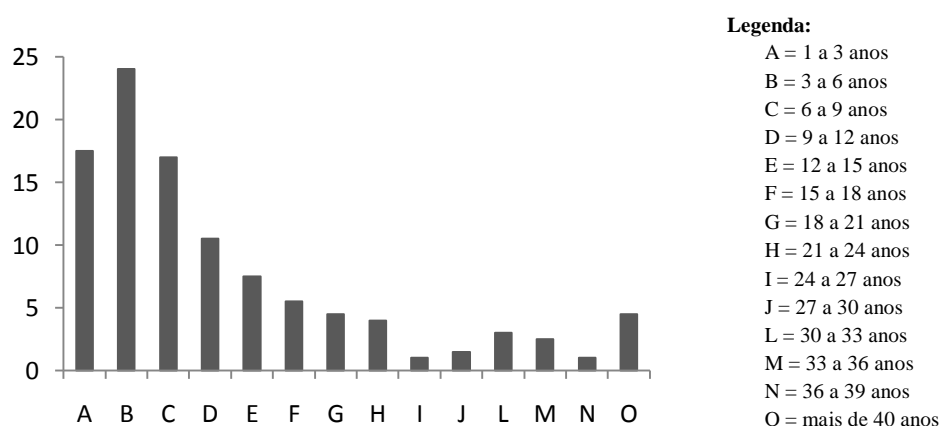


GRÁFICO 2.1: REPRESENTAÇÃO DAS IDADES DAS EMPRESAS INSOLVENTES
FONTE: Silva (2003)

Apesar dos dados apresentados no Gráfico 2.1, não se pode negar que empresas antigas e tradicionais possam deixar de pagar suas dívidas, tornando-se insolventes, nem que empresas novas não possam honrar seus compromissos, sendo adimplentes.

Outras formas de análise de capacidade poderão ser usadas, como, por exemplo, uma investigação sobre aspectos como: as instalações da empresa, a estrutura organizacional, o sistema de informação gerencial, o sistema de marketing da empresa, e também o grau de investimento em pesquisas para desenvolvimento de novos produtos e manutenção da lucratividade.

2.5.3 Capital

O capital refere-se à situação econômico-financeiro da empresa, no que diz respeito aos seus bens e recursos possuídos para saldar seus débitos (SILVA,2003).

O aspecto capital nas empresas tomadoras de empréstimos implica uma análise global, as chamadas Análise de Balanço e Análise Econômico-financeira (SCHRICKEL,2000).

O C de Capital compreende a situação econômica, financeira e patrimonial da empresa.

2.5.4 Condições

As condições dizem respeito ao micro e macrocenário em que o tomador de empréstimos está inserido. No caso das empresas, tal cenário é o ramo de atividade e a economia como um todo (SILVA, 2003).

A identificação de tendências setoriais, crescimento e recessão de mercados relacionados com o ramo de atividade da empresa, sua dependência do governo, do mercado externo, informações sobre concorrência do setor e políticas econômicas que possam vir a alterar as condições de comercialização de produtos relacionados com sua operacionalização, são alguns fatores que auxiliam a identificação das condições que podem ser favoráveis ou não ao retorno do crédito concedido (GUIMARÃES, 2002).

Fazem parte do macroambiente as variáveis econômicas (alteração nas taxas cambiais ou de juros, carga tributária, taxa de desemprego, entre outras), políticas (segurança pública,

privatização e estatização, empresas multinacionais) e psicossociais (moda, nível de alfabetização, hábitos culturais e religiosos e outros).

Por exemplo, a taxa de desemprego. Ao perderem seus postos de trabalho, as pessoas têm alteradas suas condições de honrar compromissos. O comércio, por sua vez, tem suas condições alteradas porque as pessoas que perderam o emprego deixarão de comprar ou, se comprarem a prazo, deixarão de pagar, por não terem receita. Finalmente, os profissionais liberais têm suas condições alteradas pela redução na demanda, provocada pela retração do mercado.

O microambiente é também chamado ambiente operacional e diz respeito aos aspectos com os quais o cliente interage diretamente, influenciando e sendo influenciado. Fazem parte dele as variáveis de mercado (consumidor e fornecedor), da concorrência e da sazonalidade, entre outras.

A concentração de um produto em poucos fornecedores pode gerar tendência à cartelização, impactando preços e alterando as condições das pessoas que não conseguirem se proteger.

Segundo Silva (2000), outro fator que pode ser importante é a sazonalidade do produto, isto é, empresas cujos mercados em que atuam dependem de épocas específicas, como é o caso daquelas que produzem sorvetes, que devem orientar sua produção para atender às épocas de maior demanda, das usinas de açúcar e das destilarias de álcool, cuja produção é desenvolvida durante cerca de seis meses, que é a época da safra de canas.

2.5.5 Colateral

O colateral, tradução do termo inglês de idêntica grafia, significa garantia. Procura averiguar a capacidade do tomador em oferecer garantias colaterais, espécie de segurança a mais necessária na concessão de crédito que atenua o risco da operação.

As garantias são divididas em reais e pessoais. A garantia real ocorre quando além da promessa de pagamento, o devedor confere ao credor um direito especial de garantia sobre uma coisa ou uma universalidade de coisas, destacando-se de seu patrimônio para garantir o cumprimento da obrigação. Como exemplos de garantias reais, têm-se: Hipoteca, Penhor (e Caução), Anticrese e Alienação Fiduciária. No caso das garantias pessoais, o credor conta com a promessa de pagamento e com a garantia comum que o patrimônio atual e futuro do

devedor lhe possa dar. São exemplos de garantias pessoais: o aval, carta de crédito e carta de fiança.

Porém, deve-se ressaltar que a garantia nunca deve ser o motivo para se efetuar o crédito, pois, sozinha, não faz com que o crédito retorne no prazo combinado.

Não se pode afirmar com precisão se os valores adquiridos pelo tomador retornarão. Porém, a análise minuciosa dos diversos fatores e do risco é um instrumento imprescindível na tomada de decisão.

Com isso, o risco de crédito serve para caracterizar os diversos fatores que poderão contribuir para que aquele que concedeu o crédito não receba do devedor o pagamento na época acordada.

As instituições financeiras são os principais agentes emprestadores em uma economia. Os recursos emprestados por estas instituições são provenientes de captações efetuadas junto ao público (correntistas ou poupadores) ou junto a outras instituições chamados parceiros. Assim, da mesma forma que ao conceder crédito avaliam seu risco como prestador, as instituições também são avaliadas por outras instituições quando a estas recorrem para captar recursos.

Cada instituição se utiliza de um sistema para avaliar o risco de crédito, e sua importância decorre da dificuldade que esta enfrenta em atender as solicitações de crédito e qual delas deve ser atendida considerando o retorno e o risco inerente.

Apesar do sistema de classificação apresentar imperfeições, ele se faz necessário para preservar os interesses do prestador. No próximo item serão desenvolvidos alguns exemplos de modelos e estudos a respeito de análise de crédito.

2.6 MODELOS DE PREVISÃO DE RISCO DE CRÉDITO

Com o rápido desenvolvimento da informática, a partir dos anos 70, os sistemas de pontuação de crédito baseados na abordagem estatística surgiram no negócio de financiamento de pessoas físicas e jurídicas. Tais sistemas são tidos como métodos mais importantes de tomadas de decisão para grandes volumes de solicitações de crédito (SANTOS, 2000).

Os modelos de previsão de risco de crédito vêm como uma ferramenta de auxílio massificado, que é caracterizado pela avaliação de um grande número de solicitações de pequenos valores, já que a competitividade do mercado exige decisões rápidas. O analista

informa os dados de seu potencial cliente no sistema de crédito e, imediatamente, o computador fornece a informação quanto à aprovação do crédito. O método estatístico utilizado para a construção do modelo leva em conta o histórico da instituição com seus clientes, possibilitando a identificação das características capazes de diferenciar o bom do mau pagador (SILVA, 2006).

O mundo dos negócios caminha cada vez mais em direção a um maior aprofundamento no conhecimento das atividades de seus clientes e isto faz com que métodos quantitativos assumam um papel de instrumentos auxiliares. Conforme o perfil do cliente, os modelos podem ter uma importância maior ou menor. O uso de métodos quantitativos prestará uma importante contribuição ao crédito, porém não eliminará a necessidade de que os profissionais dos bancos sejam bem treinados em créditos e negócios bancários.

Na mensuração do risco de crédito, segundo Caouette *et al* (2000), os modelos de pontuação podem ser classificados de acordo com três dimensões diferentes: as técnicas empregadas, o domínio de aplicações no processo de crédito e os produtos a que se aplicam.

As técnicas mais comumente empregadas são:

- Técnicas Econométricas: métodos estatísticos, tais como regressão logística, análise discriminante, análises *logit* e *probit* modelam a probabilidade de inadimplência como variável dependente, explicada através de um conjunto de variáveis independentes;
- Redes Neurais: sistemas computacionais que tentam simular o funcionamento do cérebro humano por uma rede de neurônios interligados. Uma alternativa às tradicionais técnicas estatísticas, por buscar a solução através de um processo de aprendizagem;
- Modelos de Otimização: técnicas de programação matemática que identifica os pesos ideais das variáveis, de forma a minimizar o erro do credor e maximizar seus lucros;
- Sistemas Especialistas: sistemas que imitam o processo usado por um analista experiente, que baseado num conjunto de regras decisórias, tentam reproduzir seu sistema estruturado, auxiliando, portanto, na tomada da decisão quanto à concessão do crédito;
- Sistemas Híbridos: sistemas movidos em parte por uma relação causal, cujos parâmetros são determinados por meio de técnicas de estimativa, utilizando computação, estimativas e simulações diretas.

2.7 REVISÃO BIBLIOGRÁFICA DAS APLICAÇÕES DOS MODELOS DE PREVISÃO

Nesta seção, são apresentados alguns estudos existentes na bibliografia a respeito dos modelos de previsão de crédito para pessoas físicas e jurídicas, utilizando estas diferentes técnicas.

Nos Estados Unidos, Altman *et al* (1968) fez uso da metodologia análise de discriminante para previsão de insolvência e classificação de empresas. Utilizou uma amostra de 58 empresas e os autores conseguiram acertos de 88% para as empresas consideradas boas e 87% para as empresas consideradas ruins.

No Brasil, Kanitz (1974) foi o pioneiro no uso de análise discriminante e construiu o chamado termômetro de insolvência. Elizabetsky (1976) desenvolveu um modelo matemático para decisão de crédito no banco comercial, usando análise discriminante, com resultado de 88,89% de acerto do modelo, em empresas classificadas como “boas”.

Steiner *et al* (1999) utilizou sistemas especialistas probabilísticos e redes neurais, ambas na área de Inteligência Artificial, na análise de crédito bancário a pessoas físicas, com o objetivo de prever o comportamento de futuros clientes à inadimplência. Tendo por base dados históricos de 2.855 clientes de um banco alemão, foram obtidos padrões, mostrando, para diagnósticos futuros, em qual situação é interessante para o banco fazer a concessão do crédito.

Guimarães e Chaves Neto (2002) utilizaram uma amostra de dados de 707 observações, sendo 102 pertencentes ao grupo de clientes inadimplentes e 605 pertencentes ao grupo de clientes adimplentes. Foram estudadas as variáveis constantes em formulários de adesão utilizados pela instituição financeira fornecedora dos dados. Algumas foram excluídas, que de alguma forma possibilitassem a identificação dos clientes, e outras foram descartadas em função do baixo poder discriminante, restando ao final 18 variáveis. O trabalho apresentou a construção e avaliação de regras de reconhecimento de padrões baseadas em duas técnicas estatísticas multivariadas: a Função Discriminante Linear de Fisher e a Regressão Logística.

O cálculo dos estimadores de máxima verossimilhança em ambas as regras foi efetuado através de um programa computacional escrito em *Visual Basic v.5.0*. Os resultados alcançados pelos dois modelos foram de 99,02% para a regressão logística e de 92,16 % para a análise discriminante.

Lemos (2003) analisou os dados históricos de 339 clientes (pessoas jurídicas) de uma agência bancária, dos quais 266 eram adimplentes e 73 inadimplentes, por meio de duas das

ferramentas de *Data Mining*³: Redes Neurais e Árvores de Decisão. De cada um dos registros foram extraídas 24 informações (variáveis). Na implementação da técnica Árvores de Decisão utilizou o *software* computacional WEKA (*Waikato Environment for Knowledge Analysis*). Os dados foram divididos em conjuntos de treinamento e teste, obtendo para o modelo 71,84% na precisão de classificação das empresas. Na implementação da técnica Redes Neurais utilizou o *software* MATLAB (*Neural Networks Toolbox*). Os treinamentos foram feitos por meio de uma rede de múltiplas camadas, usando o algoritmo *backpropagation* padrão variando parâmetros, obtendo assim para o modelo 90,04% na precisão de classificação.

Os resultados sugerem que o desempenho das Redes Neurais foi melhor que o apresentado na técnica Árvores de Decisão em relação à taxa de classificação correta, porém, do ponto de vista do usuário (analista de crédito ou gerente bancário), sempre há vantagens do uso de Árvores de Decisão no sentido de que ela apresenta resultados (regras de decisão) de fácil compreensão. Os resultados fornecidos pelos dois métodos comprovaram a eficiência na classificação de empresas adimplentes e inadimplentes (LEMOS, 2003).

Huang *et al* (2003) apresentou em seu artigo uma comparação entre a metodologia redes neurais, regressão logística e *Support Vector Machine* (SVM's). O modelo inicial tinha vinte e uma variáveis e quatro bases diferentes de dados de dois países distintos: Estados Unidos e Taiwan. O experimento foi baseado em dois conjuntos: o de treinamento e o de teste. Para a rede neural foi usado o *backpropagation* e para a avaliação do desempenho o *10-fold-cross*. A regressão logística teve uma média de 73,92% de acerto, 78,90% para SVM e 77,65% para as redes neurais. Portanto, nesse artigo, o SVM teve um melhor desempenho entre as técnicas escolhidas. A análise realizada foi em relação a instituições financeiras que trabalham com concessão de crédito, e não especificamente análise de crédito bancário. O público alvo foram grandes empresas, que para solicitar o crédito necessitavam fornecer os últimos cinco relatórios anuais e demonstrações financeiras, tal como o total de ativos e passivos circulantes, dívidas da empresa, fluxos de caixa, entre outras, que eram as variáveis do problema, e que diferem das utilizadas neste trabalho.

Gonçalves (2005) comparou regressão logística, redes neurais e algoritmos genéticos para a construção dos modelos de score de crédito. A amostra era de 10.000 clientes adimplentes e 10.000 clientes inadimplentes e divididas em: conjunto de treinamento, validação e teste, e estes dados contavam com 28 variáveis. Para a análise descritiva, categorização dos dados e aplicação de regressão logística, foi utilizado o *software* SPSS

³ Também conhecida como Mineração de Dados a qual refere-se à aplicação de algoritmos para extrair modelos dos dados.

(*Statistical Package for Social Science*) for Windows v.11.0; para a seleção das amostras e aplicação da rede neural foi utilizado o *software Enterprise Miner* v.4.1; para o algoritmo genético o autor desenvolveu um programa em *Visual Basic*. O autor avaliou o percentual de acerto total dos modelos e neste aspecto os percentuais foram muito similares nos modelos de regressão logística (68,3%) e rede neural (67,7%), e foram um pouco inferiores para o modelo de algoritmo genético (66,5%).

O Quadro 2.1, extraído do trabalho de Gonçalves (2005), mostra a variedade de resultados obtidos em outros trabalhos.

OUADRO 2.1: PRECISÃO DA CLASSIFICAÇÃO DOS MODELOS

	Regressão Linear	Regressão Logística	Árvores de Classificação	Programação Linear	Redes Neurais	Algoritmos Genéticos	Análise Discriminante	REAL ⁴
Fritz e Hoseman (2000)			79,50		81,60	82,40	82,70	
Arraes <i>et al</i> (1999)		84,80			85,40			
Chen <i>et al</i> (2002)					91,90	92,90		
Nanda e Pendharkar (2001)						65,00	62,50	
Ohtoshi (2003)		83,50	73,90		85,00			83,1
Picinini <i>et al</i> (2003)		63,50			64,40	67,50		
Arminger <i>et al</i> (1997)		67,60	66,40		65,20			
Huang <i>et al</i> (2004)		77,00			80,00			
Semolini (2002)		68,30			67,40			
Rosa (2000)		70,40	66,60					71,4

FONTE: Gonçalves (2005)

Selau (2008) apresentou um estudo sobre a construção de modelos de previsão de risco de crédito. A amostra foi de 11.681 clientes e foram consideradas 16 variáveis, que são as informações disponíveis no sistema da empresa e que o cliente preenche no momento da solicitação do crédito. Os modelos escolhidos pela autora foram: Análise discriminante, Regressão Logística e Redes Neurais. Para a construção dos modelos com uso das técnicas de análise discriminante e regressão logística, foi utilizado o *SPSS v.13.0 (Statistical Package for Social Science)*. O *software* utilizado para o treinamento e teste das redes neurais foi o *BrainMaker Professional v.3.7*. A função de ativação foi a sigmóide e o algoritmo de

⁴ A metodologia REAL (*Real Attribute Learning Algorithm*), apresentada no Quadro 1 é um modelo similar a uma árvore de classificação proposta por Steiner *et al* (1998); mais detalhes podem ser encontrados em Rosa (2000) e Ohtoshi (2003).

aprendizado supervisionado de *backpropagation* de erro, com somente uma camada escondida. Obteve resultados superiores com a regressão logística quando comparadas as outras duas técnicas empregadas.

O Quadro 2.2, construído a partir da literatura pesquisada, é similar ao Quadro 2.1 anterior (GONÇALVES, 2005), e reforça a grande variedade de resultados encontrados. Note-se que, ao se analisar os dois quadros, os modelos apresentam uma precisão de classificação entre 62,5% e 99,02%.

QUADRO 2.2: PORCENTAGEM DE PRECISÃO NA CLASSIFICAÇÃO DOS MODELOS CONSTRUÍDOS (LITERATURA PESQUISADA)

	Regressão Linear	Regressão Logística	Árvores de Classificação	Programação Linear	Redes Neurais	Algoritmos Genéticos	Análise Discriminante	SVM
Lai <i>et al</i> (2006)		75,82			80,77			79,91
Huang <i>et al</i> (2003)		73,92			77,65			78,90
Lemos <i>et al</i> (2005)			71,87		90,04			
Mendes Filho <i>et al</i> (1996)					96,60			
Guimarães (2002)		99,02					92,16	
Gonçalves (2005)		68,30			67,70	66,50		
Selau (2008)		84,10			83,50		83,90	
Velasco <i>et al</i> (2006)	67,77	68,12			68,99			
Protil <i>et al</i> (2006)		91,20					88,20	
Yin (2005)		91,00			94,00		86,00	
Eifert (2003)		88,20						
Ramirez <i>et al</i> (2007)					94,19			
Pires (2008)					96,40			

FONTE: A autora (2009)

São diversas as técnicas a serem utilizadas para a construção dos modelos de riscos de crédito. Dentre os métodos possíveis estão: regressão linear, regressão logística, árvores de classificação, programação linear, redes neurais, algoritmos genéticos, análise de discriminante e *Support Vector Machines* (SVM's).

No próximo Capítulo, serão descritas as três técnicas utilizadas neste trabalho e também o algoritmo *Sequential Minimal Optimization* (SMO), pois o modelo de otimização gerado pelo SVM envolve em sua formulação um problema quadrático. A desvantagem do problema quadrático é necessitar de ferramentas complexas para sua resolução. Em virtude

dessa dificuldade, Ales (2008) optou em utilizar em seu programa o algoritmo SMO que possibilita encontrar a solução com maior agilidade, e para isso necessitou conhecer matematicamente o desenvolvimento desse algoritmo.

CAPÍTULO III

3 TÉCNICAS UTILIZADAS

3.1 INTRODUÇÃO

Para classificar padrões, existem alguns desafios propostos ao algoritmo, podendo-se citar a robustez e a estabilidade estatística como os mais importantes (SHAW-TAYLOR e CRISTIANINI, 2000).

Existem vários paradigmas de classificação, cada um com suas vantagens e desvantagens, facilidades e limitações, sendo praticamente impossível concluir que existe um melhor classificador para todos os possíveis conjuntos de treinamento.

Nos diversos métodos de classificação, pesquisadores se inspiraram em sistemas biológicos. Seguindo essa linha, temos as Redes Neurais Artificiais (RNAs), cujos conceitos são baseados nos mecanismos de aprendizagens cerebrais, e os Algoritmos Genéticos, que têm seus conceitos inspirados no processo da evolução natural e genética. No campo do aprendizado simbólico, tem-se as Árvores de Decisão. As SVM's, por sua vez, fundamentam-se em Teorias Estatísticas.

Estas metodologias têm sido aplicadas em diversas áreas, como reconhecimento automático e contagem de partículas de matéria descritas através do tamanho e da forma (células, bactérias, viroses), identificação de impressões digitais, detecção e reconhecimento de faces, reconhecimento de textos e assinaturas, Análise de Crédito, Diagnóstico Médico, Identificação de Recursos Naturais, entre outros.

Este capítulo contém as técnicas de Classificação efetivamente utilizadas neste trabalho, com um detalhamento maior de cada técnica, os modelos matemáticos e os erros existentes na aplicação das técnicas na resolução dos problemas.

As técnicas de Classificação foram escolhidas conforme a necessidade de atingir os objetivos previstos no capítulo I.

Portanto, a predição e classificação são as tarefas a serem realizadas. Após análise de diversas técnicas de predição estudadas na literatura, optou-se por três delas, por se adequarem melhor ao banco de dados disponíveis e oferecerem uma margem de erro

aceitável, além de propiciarem rápido processamento computacional e leitura de dados acessíveis para usuários leigos. Desta forma as técnicas selecionadas foram às seguintes:

- Regressão Logística;
- Redes Neurais Artificiais (RNA);
- *Support Vector Machine (SVM)*.

Seguem detalhamentos de cada técnica.

3.2 REGRESSÃO LOGÍSTICA

3.2.1 Modelos e Técnicas Estatísticas

Ao se estudar as relações entre um conjunto de variáveis independentes com uma variável dependente, se está considerando um problema multivariado. Na análise de tal problema, geralmente usa-se algum modelo matemático para lidar com as complexas inter-relações entre as diversas variáveis.

Servigny e Renault (2004, p.107) apresentam vários fatores que são importantes na escolha de uma classe particular de modelo de mensuração de risco de crédito, em detrimento a uma outra classe:

- *Desempenho* – o fator primário. Um modelo mais complexo deveria fornecer detecção ou classificação de empresas arriscadas significativamente melhor se comparado a uma regra simples. Definir classificação ou previsão como o objetivo terá impacto na escolha.
- *Disponibilidade de dados e qualidade*. Muitos modelos têm bom desempenho em laboratório, mas não suportam dificuldades práticas, tais como valores faltantes ou *outliers*. Um modelo mais simples, mais robusto pode ser preferível a um sistema “estado da arte” em conjunto de dados imperfeitos.
- *Entendimento dos usuários*. Usuários dos modelos de escore deveriam entender perfeitamente como o modelo funciona e o que norteia os resultados. Do contrário, eles não serão capazes de reconhecer viés sistemático ou entender as limitações do seu modelo.
- *A robustez do modelo para novos dados*. Alguns modelos simulados em um dado conjunto de dados podem gerar diferentes resultados se o conjunto de dados for aumentado. Essa instabilidade deve ser evitada, pois pode indicar que o modelo está detectando o ótimo local em vez do ótimo global.

- *O tempo requerido para calibrar ou recalibrar o modelo.* Isso também importa, dependendo da frequência que o usuário quer usar para rodar o modelo⁵.

Na literatura existente sobre modelos de previsão de crédito, destacam-se, pela grande predominância, os modelos que utilizam a análise discriminante e a regressão logística. No entanto, em função da grande importância que o assunto tem adquirido e também pelo grande avanço computacional, novos modelos mais sofisticados tem surgido. Pode-se citar: Redes Neurais, Árvores de decisão, Algoritmos Genéticos e *Support Vector Machine* (SVM's).

Neste trabalho, a técnica estatística utilizada é a regressão logística. E, quando se trata de regressão, dois aspectos devem ser esclarecidos. Um deles é apresentado por Gujarati (2000, p.8):

Embora a análise de regressão lide com a dependência de uma variável em relação a outras variáveis, ela não implica necessariamente causalção. Nas palavras de Kendall e Stuart, “uma relação estatística, por mais forte e sugestiva que seja, jamais pode estabelecer uma relação causal: nossas idéias sobre causalção devem vir de fora da estatística, enfim, de outra teoria.”

Essa diferença deve ser considerada para que não se tirem conclusões equivocadas dos resultados da regressão, ou seja, as variáveis explicativas não são necessariamente as causadoras dos efeitos na variável dependente.

O outro aspecto refere-se à correlação: “Intimamente relacionada, porém conceitualmente muito diferente da análise de regressão, é a análise de correlação, cujo objetivo básico é medir a intensidade ou o grau de associação linear entre duas variáveis.” (GUJARATI, 2000). Ou seja, na regressão não se está necessariamente interessado na medida do grau de associação entre as variáveis, mas na estimativa ou previsão do valor médio de uma variável (dependente) em função dos valores fixados de outras variáveis (independentes).

5 “

- *Performance – the primary factor. A more complex modelo should significantly improved detection or classification of risky firms over a naive rule. Defining classification or prediction as the goal will have an impact on the choice.*
- *Data availability and quality. Many models perform well in the laboratory but do not cope with practical difficulties such as missing values or outliers. A simpler but more robust model may be preferable to a state-of-the-art system on patchy data sets.*
- *Understandingby users. Users of the scoring models should understand perfectly how the model works and what drives the results. Otherwise they will not be able to spot systematic bias or to understand the limits of their model.*
- *The robustness of the model to new data. Some models trained on a given data ser will provide very different results if the data set is increased slightly. This instability should be avoided, as it may mean that the model is detecting local optima rather than global optima.*
- *The time required to calibrate or recalibrate the model. This also matters, depending on what frequency the user wants to use to run the model.”*

3.2.2 Análise Discriminante x Regressão Logística

Conforme Hair Jr. *et al* (2005), na tentativa de se escolher uma técnica analítica, pode-se deparar com um problema que envolva uma variável dependente categórica e diversas variáveis independentes métricas. Utilizando o exemplo de distinção entre risco de crédito bom ou ruim, se o risco de crédito for uma medida métrica, pode-se utilizar a regressão linear múltipla. Mas, se a medida não for métrica, ou seja, se somente for possível afirmar que o risco de crédito é bom ou ruim (se pertence a um determinado grupo), não é possível utilizar a regressão linear (pois a variável dependente deve ser métrica). Quando a variável dependente é categórica (nominal ou não métrica), a análise de discriminante e a regressão logística são as técnicas estatísticas apropriadas. No entanto, a análise discriminante é capaz de trabalhar com dois ou mais grupos, mas a regressão logística é limitada, na sua forma básica⁶, a dois grupos.

Os autores (HAIR JR. *et al.*, 2005) explicam que, mesmo quando a variável dependente possui apenas dois grupos, a regressão logística pode ser preferida à análise discriminante por dois motivos:

- A análise discriminante depende do atendimento rígido dos pressupostos de normalidade multivariada e igualdade das matrizes de variância-covariância dos grupos – pressupostos que não são atendidos em muitas situações. A regressão logística não possui essas restrições e é muito mais robusta quando esses pressupostos não são atendidos, tornando-a apropriada em muitas situações.
- Mesmo quando os pressupostos são atendidos, alguns pesquisadores preferem a regressão logística porque ela é similar à regressão linear múltipla. Ela tem testes estatísticos simples, abordagens similares para incorporar variáveis métricas e não métricas e efeitos não-lineares, e uma ampla variedade de diagnósticos.⁷

Por essas razões técnicas, a regressão logística é equivalente à análise discriminante de dois grupos e pode ser mais adequada em várias situações.

⁶ Segundo os autores, outras formulações de regressão logística podem trabalhar com mais de dois grupos (categorias). Hosmer e Lemeshow (1989) apresentam o modelo de regressão linear multinomial (onde a variável dependente é nominal e possui mais de dois níveis).

⁷

- *Discriminant analysis relies on strictly meeting the assumptions of multivariate normality and equal variance-covariance matrices across groups – assumptions that are not met in many situations. Logistic regression does not face these strict assumptions and is much more robust when these assumptions are not met, making its application appropriate in many situations.*
- *Even if the assumptions are met, many researchers prefer logistic regression because it is similar to multiple regression. It has straightforward statistical approaches to incorporating metric and nonmetric variables and nonlinear effects, and a wide range of diagnostics.”*

Nesse sentido, no estudo de Wesrgaard e Wijst (2001), afirma-se que não seria possível utilizar a análise discriminante, já que ela não gera probabilidade como resultado. Assim, a análise logística e a *probit* seriam as candidatas mais óbvias para o modelo.

Especificamente para o setor bancário, Bandyopadhyay (2006) apresenta a regressão logística como solução para os bancos que pretendem estimar a PD (probabilidade dicotômica) diretamente, pois, diferentemente da análise discriminante, além de produzir probabilidade como resultado, tem a flexibilidade de incorporar variáveis financeiras e não financeiras.

3.2.3 Regressão Logística

Nos modelos de regressão logística, a variável dependente é, em geral, uma variável binária (nominal ou ordinal) e as variáveis independentes podem ser categóricas (desde que dicotomizadas após transformação) ou contínuas.

Considere o caso em que as observações podem ser classificadas em uma de duas categorias mutuamente exclusivas (0 e 1). Como exemplo, as categorias poderiam representar uma empresa que pode ser classificada como adimplente ou inadimplente.

A variável dependente binária Y pode assumir valores:

$$Y_i = \begin{cases} 1 & \text{Se o } i - \text{ésimo indivíduo pertencer a categoria dos adimplentes} \\ 0 & \text{Se o } i - \text{ésimo indivíduo pertencer a categoria dos inadimplentes} \end{cases}$$

O modelo de regressão logística é também conhecido como modelo logístico e é baseado na função sigmóide, $f(z)$, dada por:

$$f(z) = \frac{1}{1+e^{-z}} \quad (3.1)$$

ou ainda, habitualmente mais usada

$$f(z) = \frac{e^z}{1+e^z} \quad (3.2)$$

Portanto, pode-se observar que a função logística varia entre 0 e 1, $0 \leq f(z) \leq 1$ e essa é a principal razão do modelo logístico ser usado para descrever uma dicotômica probabilidade de algo acontecer ou não acontecer, ou seja, variável dicotômica.

Existem ainda outras duas características de $f(z)$ que tornam o modelo logístico largamente utilizado:

- z representa um índice que combina a contribuição de diversos fatores de risco, e $f(z)$ representa o risco (probabilidade) de que um evento ocorra, para um dado z ;
- possui forma de “S”, indicando que o efeito de z em $f(z)$ é mínimo até que algum “gatilho” seja disparado, depois aumenta rapidamente até que algum nível seja alcançado, voltando a crescer lentamente.

Segue o gráfico 3.1 que representa esta função:

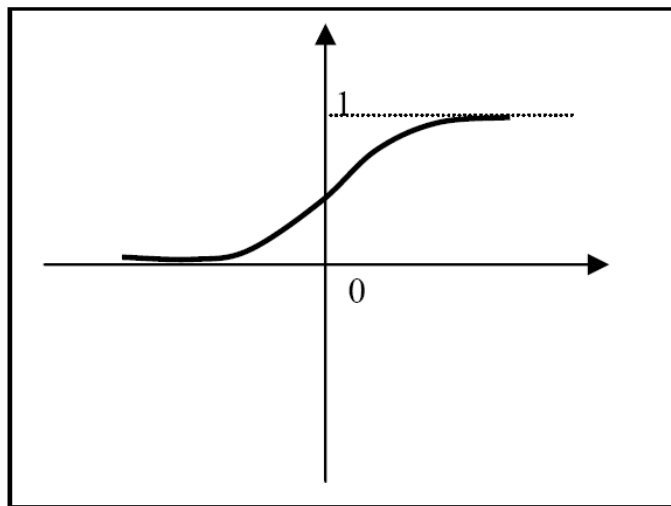


GRÁFICO 3.1: FUNÇÃO LOGÍSTICA OU SIGMÓIDE
FONTE: Adaptado de Cunico (2005)

3.2.4 O Modelo Matemático Logístico

A partir da função logística $f(z)$, pode-se obter o modelo logístico, escrevendo z como a soma linear das variáveis independentes e substituindo na função:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (3.3)$$

Assim, a função logística passa a ser representada como a contribuição das variáveis independentes x e os parâmetros desconhecidos $\beta_0, \beta_1, \beta_2, \dots, \beta_n$. Para que ela torne-se o

modelo logístico, deve-se escrever na forma específica, substituindo $f(z)$, por $\pi(x)$ como segue:

$$E(Y|x) = \pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}} \quad (3.4)$$

onde $E(Y|x)$ é a esperança de Y dado x .

3.2.5 Transformação Logit

A Transformação Logit ou logito é interessante e necessária pois estabelece uma relação linear entre as variáveis explicativas e a transformada da variável resposta. A demonstração da transformação segue abaixo:

Partindo-se da equação

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}} \quad (3.5)$$

chamando $g = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$

e resolvendo para $g(x)$, tem-se:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] \quad (3.6)$$

sendo que a função logit $g(x)$ é agora linear nos seus parâmetros.

3.2.6 Ajuste do Modelo de Regressão Logística

O ajuste pode ser realizado através do método dos mínimos quadrados não linear ou pelo método da máxima verossimilhança (HOSMER e LEMESHOW, 1989). O método de máxima verossimilhança é sempre mais indicado por possuir propriedades ótimas⁸. Desta

⁸ As propriedades são assintóticas para estimar os parâmetros, além de serem consistentes e assintoticamente eficientes.

forma, tem-se as estimativas dos parâmetros $\beta_0, \beta_1, \dots, \beta_n$ dadas por $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_n$ que por sua vez alimentarão o modelo estimado e a logit estimada que ficam representadas na seguinte forma:

- Modelo estimado, $\hat{\pi}(x) = \frac{e^{\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \dots + \widehat{\beta}_n x_n}}{1 + e^{\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \dots + \widehat{\beta}_n x_n}}$
- Logit Estimada, $\hat{g}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \dots + \widehat{\beta}_n x_n$

Assim, estimados os valores dos coeficientes $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_n$, basta substituí-los nas funções acima para encontrar o valor estimado de $\hat{\pi}(x)$, que logicamente será um valor entre 0 e 1, ou seja, $0 \leq \hat{\pi}(x) \leq 1$. De acordo com esse valor, classifica-se o novo indivíduo na respectiva categoria dicotômica.

3.2.7 Verificação do Ajuste

A qualidade do modelo ajustado através da regressão logística pode ser verificada através da comparação entre os valores observados e os valores preditos para a variável resposta. De forma bastante simples, constrói-se uma tabela com os valores observados versus os valores preditos e verifica-se a existência de uma relação linear entre essas duas quantidades, conforme segue no quadro 3.1, elaborado para um exemplo de 200 indivíduos classificados se são portadores de celulares (1) ou não portadores (0).

QUADRO 3.1: EXEMPLO DE CLASSIFICAÇÃO DOS 200 CASOS DE PORTADORES E NÃO PORTADORES DE CELULAR

Variável dicotômica	Preditas em 0	Preditas em 1	% de acerto
0	55	15	78,57%
1	5	125	96,15%

FONTE: A autora (2009)

3.3 REDES NEURAIIS

3.3.1 Aprendizagem de máquina

A Inteligência Artificial (IA) é construída a partir dos fundamentos da heurística, em oposto à Estatística, tenta imitar a maneira como o homem pensa na resolução dos problemas estatísticos.

Segundo Monard and Baranauskas (2003), Aprendizagem de Máquina (AM) é uma área da Inteligência Artificial (IA) cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado, bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseados em experiências acumuladas através da solução bem sucedida de problemas anteriores (MONARD e BARANAUSKAS, 2003).

Aprendizagem de máquina, também chamada de *machine learning*, é um termo que engloba um conjunto de metodologias e comportamentos em dados que representam exemplos de acontecimentos do mundo real ou experiências passadas. Dessa forma, os dois objetivos de qualquer projeto de aprendizagem são: induzir o modelo processando uma grande quantidade de dados e realizar inferências a partir dele. Dentre esses objetivos, processar essa grande quantidade de dados é a que exige maior tempo e esforço computacional.

O aprendizado indutivo é efetuado a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema de aprendizado. O aprendizado indutivo pode ser dividido em dois tipos principais: supervisionado e não-supervisionado.

No aprendizado supervisionado, é fornecido ao algoritmo de aprendizagem, um conjunto de exemplos de treinamento na forma: entrada, saída desejada. O objetivo do algoritmo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não-rotulados, ou seja, exemplos que não tenham o rótulo da classe (saída conhecida). Para rótulos de classes discretos é conhecido como classificação e para valores contínuos como regressão.

No aprendizado não-supervisionado, para cada exemplo apenas os atributos de entrada estão disponíveis. O indutor analisa os exemplos fornecidos e tenta determinar se alguns deles

podem ser agrupados de alguma maneira, formando os agrupamentos ou *clusters* (CHEESEMAN e STUSTZ 1990).

Na Figura 3.1 é mostrada a hierarquia do aprendizado descrita acima:



FIGURA 3.1: HIERARQUIA DO APRENDIZADO
FONTE: A Autora (2009)

Um requisito importante para as técnicas de aprendizagem de máquina é que elas sejam capazes de lidar com dados imperfeitos, denominados ruídos, ou seja, valores afetados por leituras de medidas incorretas (SHAWER-TAYLOR e CRISTIANINI, 2000).

3.3.2 Uma introdução às Redes Neurais

O cérebro humano é considerado o mais fascinante processador baseado em carbono existente, sendo composto por aproximadamente 10 bilhões de neurônios (CARVALHO, S.d.). Todas as funções e movimentos do organismo estão relacionados ao funcionamento destas pequenas células. Os neurônios estão conectados uns aos outros através de sinapses, e juntos formam uma grande rede, chamada REDE NEURAL. As sinapses transmitem estímulos através de diferentes concentrações de Na⁺ (Sódio) e K⁺ (Potássio), e o resultado disto pode ser estendido por todo o corpo humano. Esta grande rede proporciona uma fabulosa capacidade de processamento e armazenamento de informação.

O sistema nervoso é formado por um conjunto extremamente complexo de neurônios. Nos neurônios a comunicação é realizada através de impulsos. Quando um impulso é recebido, o neurônio o processa, e passado um limite de ação, dispara um segundo impulso, que produz uma substância neurotransmissora, a qual flui do corpo celular para o axônio (que por sua vez pode ou não estar conectado a um dendrito de outra célula). O neurônio que transmite o pulso pode controlar a frequência de pulsos aumentando ou diminuindo a

polaridade na membrana pós-sináptica. Ele tem um papel essencial na determinação do funcionamento, do comportamento e do raciocínio do ser humano. Ao contrário das redes neurais artificiais, redes neurais naturais não transmitem sinais negativos, sua ativação é medida pela frequência com que emite pulsos, frequência esta de pulsos contínuos e positivos. As redes naturais não são uniformes como as redes artificiais, e apresentam uniformidade apenas em alguns pontos do organismo. Seus pulsos não são síncronos ou assíncronos, devido ao fato de não serem contínuos, o que a difere de redes artificiais.

Os principais componentes dos neurônios são:

- Os dendritos, que tem por função, receber os estímulos transmitidos pelos outros neurônios;
- O corpo de neurônio, também chamado de soma, que é responsável por coletar e combinar informações vindas de outros neurônios;
- E finalmente o axônio, que é constituído de uma fibra tubular que pode alcançar até alguns metros, e é responsável por transmitir os estímulos para outras células.

A Figura 3.2 ilustra os principais componentes do neurônio:

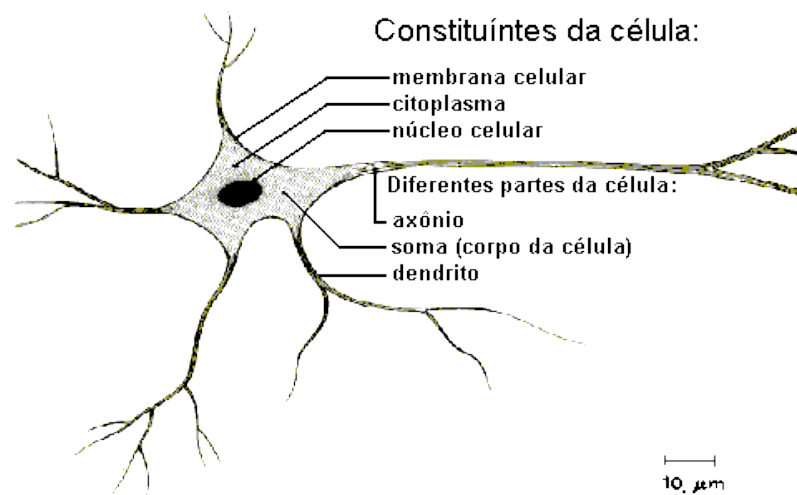


FIGURA 3.2: CONSTITUINTES DA CÉLULA NEURAL
FONTE: Tatibana (2000)

3.3.3 Histórico das Redes Neurais Artificiais

No Quadro 3.2, se encontra um resumo dos fatos históricos das Redes Neurais Artificiais em ordem cronológica:

QUADRO 3.2: HISTÓRICO DAS REDES NEURAIS ARTIFICIAIS EM ORDEM CRONOLÓGICA

Período	Fato
1943	McCULLOUGH e PITTS estabeleceram as bases da neurocomputação, com modelos matemáticos.
1949	HEBB traduziu matematicamente a sinapse dos neurônios biológicos.
1951	MINSKI constrói o Snark, primeiro neurocomputador com capacidade de aprendizado, ou seja, ajustava automaticamente os pesos entre as sinapses. Não executou nenhuma função útil.
1957	ROSENBLATT concebeu o "perceptron", que era uma rede neural de duas camadas, usado no reconhecimento de caracteres.
1958	ROSENBLATT mostrou em seu livro (<i>Principles of Neurodynamics</i>) o modelo dos "perceptrons".
1960	Surgiu a Rede Adaline e o Madaline perceptron, proposto por WIDROW e HOFF.
1962	WIDROW desenvolveu um processador para redes neurais e fundou a primeira empresa de circuitos neurais digitais, a Memistor Corporation.
1967	Fim das verbas destinadas à pesquisa de redes neurais.
1974	WERBOS lançou bases para o algoritmo de retropropagação (<i>backpropagation</i>).
1980	HOPFIELD propôs a implementação de memórias associativas por uma rede recorrente.
1986	RUMELHART, HINTON e WILLIAMS, introduziram o poderoso método <i>backpropagation</i> .

FONTE: A Autora, adaptado de Tatibana (2000).

3.3.4 Conceitos, características e estrutura da RNA

Rede neural é uma das técnicas de tratamento de dados mais recentes e que tem despertado grande interesse tanto de pesquisadores da área de tecnologia quanto da área de negócios (CORRAR *et al.*, 2007).

Rede neural artificial (RNA) pode ser definida como um método de solucionar problemas de inteligência artificial, constituída de um sistema de circuitos que simulam o cérebro humano, inclusive seu comportamento. São modelos matemáticos-computacionais inspirados no funcionamento das células neuronais, isto é, na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência (a rede aprende, corrige erros e faz descobertas com a alimentação de dados) (FAUSETT, 1994).

Assim como o cérebro humano, as redes neurais artificiais se destacam pelo alto nível de interconexão e paralelismo maciço, isto é, vários dados sendo operados simultaneamente, onde suas informações são executadas por unidades denominadas neurônios. A informação

entre os neurônios é transmitida por meio de conexões denominadas sinapses ou pesos sinápticos. Os pesos sinápticos são alterados através de processos algorítmicos, para que possam representar o conhecimento adquirido. Estas alterações caracterizam o processo de aprendizagem de uma rede neural artificial (FAUSETT, 1994).

Uma rede neural é formada por um conjunto de neurônios, também chamados de elementos de processamento. As unidades de processamento são distribuídas em um conjunto de camadas. Uma rede neural possui uma estrutura que pode ser escrita a partir de quatro conceitos (HAIR JR. *et al.*, 1998): o tipo de modelo de rede neural; as unidades de processamento que tratam a informação que entra na rede; um sistema de nós que interligam as unidades de processamento; uma função de aprendizado, a partir da qual o sistema calibra o modelo.

A propriedade mais importante das redes neurais é a habilidade de aprender de seu ambiente, e com isso melhorar seu desempenho. Isso é feito através de um processo iterativo de ajustes aplicados aos seus pesos, o treinamento.

Muitos são os modelos de redes neurais, mas os modelos básicos são: *perceptron*, Redes Lineares e Redes de Múltiplas Camadas (MLP).

O *perceptron* foi proposto por Rosenblatt em 1959 e são redes do tipo *feed-forward* constituídas de unidade binária, sendo o primeiro modelo de redes neurais. Este modelo de rede neural possui duas camadas: uma de entrada com n -elementos e uma de saída com número variável de elementos, dependendo do problema (KRÖSE e SMAGT, 1993).

As Redes Lineares diferem do *perceptron* na função de transferência que, neste caso, é linear, permitindo que as saídas tomem qualquer valor entre “0” e “1” e não apenas os valores “0” e “1” como na função passo utilizada no *perceptron*. Estas redes utilizam a regra de aprendizagem de *Widrow-Hoff*, também conhecida como regra dos Mínimos Quadrados, para ajustar os pesos e *biases* de acordo com a magnitude dos erros, e não apenas pela sua presença (STEINER, 1999).

As Redes de Múltiplas Camadas são modelos de redes que apresentam uma ou mais camadas de entrada e saída, chamadas camadas intermediárias. Este tipo de Rede Neural Artificial é o modelo mais utilizado atualmente; geralmente são treinadas através do algoritmo de *backpropagation*.

Os vários processos de aprendizagem de uma RNA podem ser classificados em:

- Aprendizado supervisionado: ou aprendizado por correção de erro. Normalmente, neste aprendizado, são fornecidas referências do objetivo a ser alcançado, com dados de entrada e de saída para o processamento da rede. Este aprendizado é utilizado no modelo

Multi Layer Perceptron (MLP), com algoritmo *backpropagation*, onde o erro pode ser calculado como a diferença entre a saída real e a saída desejada.

- Aprendizado não-supervisionado: Neste aprendizado não é apresentada nenhuma referência externa, podendo ser encontrado no modelo *hopfield*.

A estrutura de uma rede neural é composta pelos fatores:

- Arquitetura;
- Topologia;
- Seleção da função de ativação dos neurônios da camada escondida e de saída;
- Transformação dos dados ou normalização;
- Seleção do conjunto de dados;
- Algoritmo de treinamento e seus respectivos parâmetros;
- Critério de parada do treinamento.

A arquitetura da RNA refere-se ao formato da rede. Toda rede é dividida em camadas, usualmente classificadas em três grupos, onde os neurônios de uma camada são conectados apenas aos neurônios da camada imediatamente posterior, não havendo conexões entre neurônios da mesma camada. Porém, caracteristicamente, as camadas são totalmente conectadas, sendo assim definidas: camada de entrada (onde os padrões são apresentados a rede), camada(s) intermediária(s) (onde é feita a maior parte do processamento, por meio das conexões ponderadas; podem ser consideradas como extratoras de características) e camada de saída (onde o resultado final é concluído e apresentado).

A Figura 3.3 ilustra a arquitetura de uma RNA:

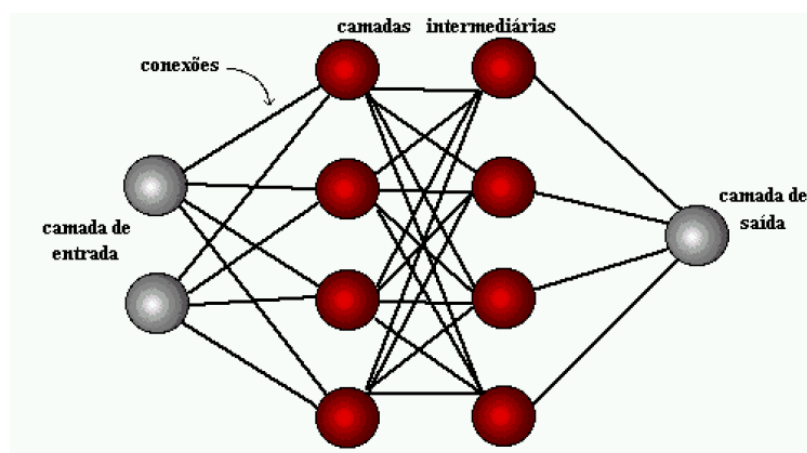


FIGURA 3.3: CONEXÕES EM UMA REDE NEURAL ARTIFICIAL
FONTE: Carvalho (S.d.)

Existem basicamente três tipos principais de arquiteturas (HAYKIN, 1999): redes *feedforward* (redes em avanço/para frente) com uma única camada, redes *feedforward* com múltiplas camadas, e redes recorrentes⁹.

Com referência à topologia da rede deve-se ter:

- o número de nós da camada de entrada: que corresponde ao número de variáveis do problema;
- o número de camadas escondidas e o número de neurônios a serem colocados nesta camada: inexistente um critério pré-determinado sobre o número de neurônios que deve ser utilizado; em geral, redes neurais com poucos neurônios são preferidas, pois tendem a possuir um maior poder de generalização, evitando, assim, o sobreajuste (*overfitting*). Porém, se a rede possui poucos neurônios, a rede não aprende, podendo ocorrer *underfitting* (a rede não converge durante o treinamento);
- o número de neurônios da camada de saída.

Acerca da função de ativação, também chamada de função de transferência, é uma função matemática que, aplicada à combinação linear entre as variáveis de entrada e os pesos que chegam a determinado neurônio, retorna ao seu valor de saída. Existem várias funções matemáticas utilizadas. Porém, comumente utiliza-se a função linear, a função logística ou a função tangente hiperbólica.

Em relação à seleção dos dados, ao implementar uma rede neural, o conjunto de dados é separado em dois subconjuntos: o de treinamento e o de teste. O conjunto de treinamento é utilizado para ajustar os parâmetros da rede, e o conjunto de teste para validar o modelo.

A um conjunto de regras bem definidas para a solução de um problema de aprendizado denomina-se algoritmo de aprendizado. Existem vários tipos de algoritmos de aprendizado, tais como, *backpropagation*, *perceptron*, entre outros, específicos para determinados modelos de redes neurais, como a MLP, Hopfield, etc. Normalmente, estes algoritmos diferem entre si, pelo modo como os pesos são modificados (TATIBANA; KAETSU, 2000).

Com relação ao algoritmo de treinamento Santos *et al.* (2005) destaca que existem vários métodos para o treinamento de uma rede neural, mas que o algoritmo mais popularmente usado é *backpropagation*. A aplicação deste algoritmo requer a escolha de um

⁹ RNAs que apresentam conexões com *loops*. São mais utilizadas em sistemas dinâmicos. Em uma rede corrente, as entradas passadas podem ser lembradas e influenciar a próxima saída gerada (MARTINELLI, 1999).

conjunto de parâmetros (número de iterações do algoritmo, critério de parada, pesos iniciais, taxa de aprendizado), cuja influência pode ser decisiva para a capacidade de generalização da rede.

Desta forma, após o treinamento, o conjunto de teste é utilizado para verificar a habilidade de generalização da rede sob condições reais de utilização. A habilidade de generalização da rede se refere ao seu desempenho ao classificar padrões do conjunto de teste (SANTOS *et al.*, 2005).

Como critério de parada do algoritmo (*backpropagation*), em geral, é utilizado um número máximo de ciclos ou iterações.

3.4 SUPPORT VECTOR MACHINES (SVM)

Support Vector Machines (SVM) são sistemas de aprendizagem de máquina que usam espaços de hipóteses de funções lineares na alta dimensionalidade característica do espaço, treinado como algoritmo de aprendizagem da Teoria de Otimização, que implementa um limite derivado da Teoria da Aprendizagem Estatística. Esta aprendizagem estratégica, introduzida por Vapnik e colaboradores, vem superando, em poucos anos desde sua introdução, outros sistemas de aprendizagem em ampla variedade de aplicações com maior desempenho (SHOWER-TAYLOR e CRISTIANINI, 2000).

Essa técnica pode ser aplicada no Reconhecimento de Padrões, Regressão linear, Extração de Características e Detecção de Novidades. Para cada um desses contextos há definições matemáticas específicas. Como o problema, investigado nesta dissertação, é de classificação e reconhecimento de empresas adimplentes e inadimplentes, encaixa-se na área de Reconhecimento de Padrões. Esta Seção fundamenta a teoria do SVM, dando ênfase a essa área.

3.4.1 Hiperplano ótimo para padrões linearmente separáveis

Considere uma amostra de treinamento $(x_i, d_i)_{i=1}^N$, onde x_i é o padrão de entrada para o i -ésimo exemplo e d_i é a resposta desejada correspondente (saída-alvo). Para começar, assumimos que o padrão (classe) representado pelo subconjunto $d_i = +1$ e o padrão

representado pelo subconjunto $d_i = -1$ são “linearmente separáveis”. A equação de uma superfície de decisão na forma de um hiperplano que realiza esta separação é

$$w^T x + b = 0 \quad (3.7)$$

onde x é um vetor de entrada, w é um vetor peso ajustável e b é um *bias*. Podemos assim escrever:

$$\begin{aligned} w^T x_i + b &\geq 0, \text{ para } d_i = +1 \\ w^T x_i + b &< 0, \text{ para } d_i = -1 \end{aligned} \quad (3.8)$$

Para um dado vetor de peso de w e *bias* b , a separação entre o hiperplano definida na equação 3.7 e o ponto de dado mais próximo é denominada *margem de separação*, representada por ρ . O objetivo de uma máquina de vetor de suporte é encontrar o hiperplano particular para o qual a *margem de separação* ρ é máxima. Sob esta condição, a superfície de decisão é referida como o hiperplano ótimo.

O modelo mais simples de SVM, que também foi o primeiro a ser introduzido, é chamado de Classificador de Margem Máxima. Ele trabalha apenas com dados linearmente separáveis, ficando restrito a poucas aplicações práticas. Apesar dessa limitação, o Classificador de Margem Máxima apresenta propriedades importantes e é a pedra fundamental para a formulação de SVMs mais sofisticadas.

A figura 3.5 ilustra a construção geométrica de um hiperplano ótimo para um espaço de entrada bidimensional. Existem vários hiperplanos que separam os dados (Figura 3.4), mas há somente um que maximiza a *margem*, ou seja, que maximiza a distância entre os pontos mais próximos de cada classe e o hiperplano de separação.

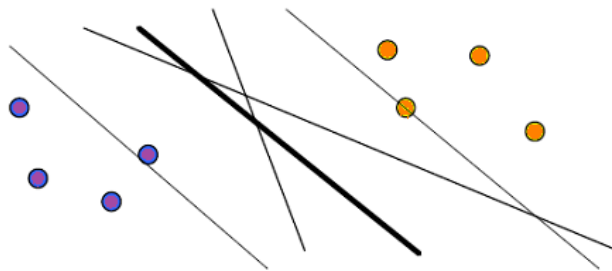


FIGURA 3.4: POSSIBILIDADES DE HIPERPLANOS SEPARADORES
FONTE: A autora (2009)

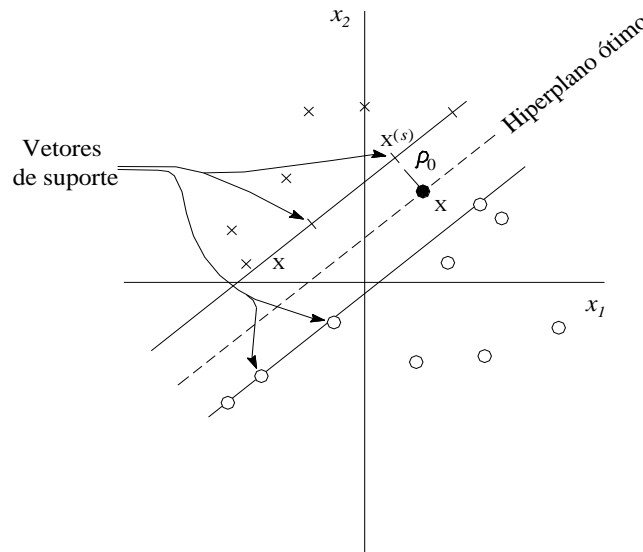


FIGURA 3.5: FORMAÇÃO DO HIPERPLANO DE SEPARAÇÃO ATRAVÉS DOS VETORES DE SUPORTE
FONTE: Haykin (1999)

Considere que w_0 e b_0 representem os valores ótimos do vetor peso e do *bias*, respectivamente. Consequentemente, o hiperplano ótimo, representando uma superfície de decisão linear multidimensional no espaço de entrada, é definido por

$$w^T x + b_0 = 0 \quad (3.9)$$

que é a Equação 3.7 reescrita. A função discriminante

$$g(x) = w_0^T x + b_0 \quad (3.10)$$

fornece uma medida algébrica da distância de x até o hiperplano (DUDA e HART, 1973). Talvez o modo mais fácil de ver isto seja expressar x como

$$x = x_p + r \frac{w_0}{\|w_0\|} \quad (3.11)$$

onde x_p é a projeção normal de x sobre o hiperplano ótimo, e r é a distância algébrica desejada; r é positivo se x estiver no lado positivo do hiperplano ótimo e negativo se x estiver no lado negativo. Como, por definição, $g(x_p) = 0$, resulta que

$$g(x) = w_0^T x + b_0 = r \|w_0\| \quad (3.12)$$

ou

$$r = \frac{g(x)}{\|w_0\|} \quad (3.13)$$

Em particular, a distância da origem (i.e., $x=0$) até o hiperplano ótimo é dada por $b_0 / \|w_0\|$. Se $b_0 > 0$, a origem está no lado positivo do hiperplano ótimo; se $b_0 < 0$, ela está no lado negativo. Se $b_0 = 0$, o hiperplano ótimo passa pela origem. Uma interpretação geométrica destes resultados algébricos é dada na figura 3.6.

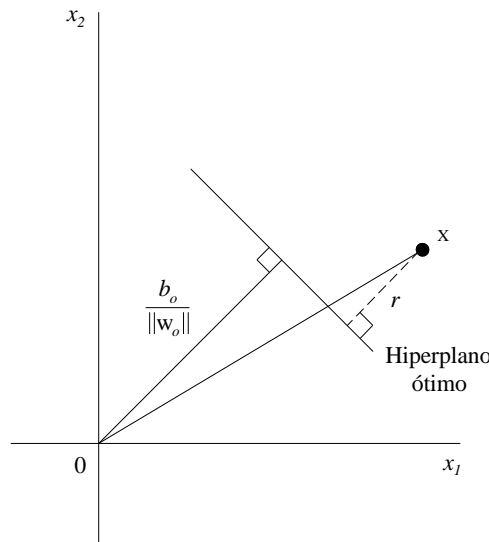


FIGURA 3.6: INTERPRETAÇÃO DAS DISTÂNCIAS ALGÉBRICAS DE PONTOS ATÉ O HIPERPLANO ÓTIMO PARA UM CASO BIDIMENSIONAL
FONTE: Haykin (1999)

O processo de treinamento do SVM consiste na obtenção de valores para os pesos w e do termo *bias* b de forma a maximizar a distância entre as margens. Desta maneira, o SVM se torna robusto a pequenas variações no conjunto de treinamento, possibilitando uma melhor generalização.

Dado o conjunto de treinamento $T = \{(x_i, d_i)\}$. Com base nos resultados retratados na figura acima, vemos que o par (w_0, b_0) deve satisfazer restrição:

$$w_0^T x_i + b \geq 1 \text{ para } d_i = +1$$

$$w_0^T x_i + b_0 < 0 \text{ para } d_i = -1 \quad (3.14)$$

Note que se a Equação 3.8 for válida, isto é, os padrões forem linearmente separáveis, podemos sempre escalar w_0 e b_0 de modo que a Equação 3.14 seja válida; esta operação de escalamento não afeta a Equação 3.9.

Os pontos de dados particulares (x_i, d_i) para os quais a primeira ou segunda linha da Equação 3.14 é satisfeita com o sinal de igualdade são chamados de *vetores de suporte*, por isso o nome “dispositivo de vetor de suporte”. Estes vetores desempenham um papel proeminente na operação desta classe de dispositivos de aprendizagem. Em termos conceituais, os vetores de suporte são aqueles pontos de dados que se encontram mais próximos da superfície de decisão e são, portanto, os mais difíceis de classificar. Dessa forma, têm uma influência direta na localização ótima da superfície de decisão.

Considere um vetor de suporte

x^s para o qual $d^s = +1$. Então, por definição, temos:

$$g(x^s) = w_0^T x^{(s)} \mp b_0 \mp 1 \text{ para } d^{(s)} = \mp 1 \quad (3.15)$$

Da Equação 3.13, a distância algébrica do vetor de suporte x^s até o hiperplano ótimo é:

$$r = \frac{g(x^s)}{\|w_0\|} = \begin{cases} \frac{1}{\|w_0\|} & \text{se } d^{(s)} = +1 \\ -\frac{1}{\|w_0\|} & \text{se } d^{(s)} = -1 \end{cases} \quad (3.16)$$

onde o sinal positivo indica que x^s se encontra no lado positivo do hiperplano ótimo e o sinal negativo indica que x^s está no lado negativo do hiperplano ótimo. Considere que ρ represente o valor ótimo da margem de separação entre as duas classes que constituem o conjunto de treinamento T. Então, da Equação 3.16 resulta que:

$$\begin{aligned}\rho &= 2r \\ &= \frac{2}{\|w_0\|}\end{aligned}\tag{3.17}$$

A equação 3.17 afirma que maximizar a margem de separação entre classes é equivalente a minimizar a norma euclidiana do vetor peso w .

Em resumo, o hiperplano ótimo definido na Equação 3.9 é *único* no sentido de que o vetor peso w_0 fornece a máxima separação possível entre exemplos positivos e negativos. Esta condição ótima é alcançada minimizando-se a norma euclidiana do vetor peso w .

3.4.2 Otimização Quadrática para Encontrar o Hiperplano Ótimo

O objetivo é desenvolver um procedimento eficiente do ponto de vista computacional para, utilizando a amostra de treinamento $T = \{(x_i, d_i)\}_{i=1}^N$, encontrar o hiperplano ótimo, sujeito à restrição

$$d_i(w^T x_i + b) \geq 1 \text{ para } i = 1, 2, \dots, N\tag{3.18}$$

Esta restrição combina as duas linhas da Equação 3.14 com w usado no lugar de w_0 .

A maximização da margem de separação entre as classes, quando os conjuntos são linearmente separáveis, com o objetivo de obter um hiperplano ótimo através do SVM, é resolvida através de um problema de otimização quadrática.

Um problema de otimização como o problema primal do SVM, pode ser descrito de forma geral:

1) Dada a amostra de treinamento $\{(x_i, d_i)\}_{i=1}^N$, encontre os valores ótimos do vetor peso w e *bias* b de modo que satisfaçam as restrições

$$d_i(w^T x_i + b) \geq 1 \text{ para } i = 1, 2, \dots, N$$

2) Vetor peso w minimize a função custo:

$$\Phi(w) = \frac{1}{2} w^T w$$

A proposta do treinamento SVM pode-se restringir para o caso onde as restrições são lineares, a função objetivo é convexa e quadrática, portanto considera-se um programa quadrático convexo (SHAWE-TAYLOR e CRISTIANINI, 2000).

Para realizar a procura de máximos e mínimos condicionados, como nos problemas quadráticos convexos e demais problemas de otimização matemática, é utilizado o método de Lagrange.

Primeiro é construída a *função lagrangiana*:

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [d_i (w^T x_i + b) - 1] \quad (3.19)$$

onde as variáveis auxiliares não-negativas α_i são chamadas de *multiplicadores de Lagrange*.

Para solucionar o problema de otimização restrito, deve-se encontrar o ponto de sela da função lagrangiana $J(w, b, \alpha)$. A derivada de $J(w, b, \alpha)$ em relação a w e b deve ser nula. Isso corresponde ao fato de que, no ponto ótimo, têm-se às seguintes equações de ponto de sela:

$$\frac{\partial J(w, b, \alpha)}{\partial w} = w - \sum_{i=1}^N \alpha_i d_i x_i = 0$$

$$\frac{\partial J(w, b, \alpha)}{\partial b} = \sum_{i=1}^N \alpha_i d_i = 0$$

Substituindo as relações obtidas, têm-se:

$$w = \sum_{i=1}^N \alpha_i d_i x_i \quad (3.20)$$

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (3.21)$$

O vetor solução w é definido em termos de uma expansão que envolve os N exemplos de treinamento. Note, entretanto, que, embora esta solução seja única em virtude da convexidade da lagrangiana, o mesmo não pode ser dito sobre os coeficientes de Lagrange, α_i .

Também é importante notar que no ponto de sela, para cada multiplicador de Lagrange α_i , o produto daquele multiplicador pela sua restrição correspondente desaparece, como mostrado por

$$\alpha_i [d_i(w^T x_i + b) - 1] = 0 \text{ para } i = 1, 2, \dots, N \quad (3.22)$$

Dessa forma, apenas aqueles multiplicadores que satisfazem exatamente a Equação 3.22 podem assumir valores *não-nulos*. Esta propriedade resulta das *condições de Karush - Kuhn-Tucker* (KKT) da teoria da otimização (BERTSEKAS, 1995).

Como notado anteriormente, o problema primal lida com uma função de custo convexa e com restrições lineares. Dado um problema de otimização restrito como este, é possível construir outro problema, chamado de *problema dual*. Este segundo problema tem o mesmo valor ótimo do problema primal, mas com os multiplicadores de Lagrange fornecendo a solução ótima. Em particular, podemos formular o seguinte *teorema da dualidade* (BERTSEKAS, 1995):

- Se o problema primal tem uma solução ótima, então o problema dual também tem uma solução ótima, e os valores ótimos correspondentes são iguais.
- Para que w seja uma solução primal ótima e α_0 seja uma solução dual ótima, é necessário e suficiente que w_0 seja realizável para o problema primal, e

$$\Phi(w_0) = J(w_0, b_0, \alpha_0) = \min_w J(w, b_0, \alpha_0)$$

Para postular o problema dual para o problema primal, primeiro expande-se a Equação 3.19, termo a termo, como segue:

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i d_i w^T x_i - b \sum_{i=1}^N \alpha_i d_i + \sum_{i=1}^N \alpha \quad (3.23)$$

O terceiro termo do lado direito da Equação 3.23 é zero em virtude da condição de otimização da Equação 3.21. Além disso, da Equação 3.20 temos

$$w^T w = \sum_{i=1}^N \alpha_i d_i w^T x_i = b \sum_{i=1}^N \sum_{j=1}^N \sum_{i=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j$$

Conseqüentemente, fazendo a função objetivo $J(w, b, \alpha) = Q(\alpha)$, pode-se reformular a Equação 3.23 como:

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j \quad (3.24)$$

onde os α_i são não-negativos.

Pode-se agora formular o problema dual:

1) Dada a amostra de treinamento $(x_i, d_i)_{i=1}^M$, encontre os multiplicadores de Lagrange α_i que maximizam a função objetivo

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j$$

2) Sujeita às restrições:

$$\begin{aligned} \sum_{i=1}^N \alpha_i d_i &= 0 \\ \alpha_i &\geq 0 \text{ para } i = 1, 2, \dots, N \end{aligned}$$

Nota-se que o problema dual é formulado inteiramente em termos dos dados de treinamento. Além disso, a função $Q(x)$ a ser maximizada depende apenas dos padrões de entrada na forma de um conjunto de produtos escalares, $\sum_{j=1}^M x_j^T x_j$.

Determinado os multiplicadores de Lagrange ótimos, representados por $\alpha_{0,i}$, pode-se calcular o vetor peso ótimo w_0 usando a Equação 3.20 e assim escrever:

$$w_0 = \sum_{i=1}^N \alpha_{0,i} d_i x_i \quad (3.25)$$

Para calcular o *bias* ótimo b_0 , pode-se usar o w_0 assim obtido e tirar vantagem da Equação 3.15 relativa ao vetor de suporte positivo, e assim escrever:

$$b_0 = 1 - w_0^T x^{(s)} \text{ para } d^{(s)} = 1 \quad (3.26)$$

O Classificador de Margem Máxima, quando aplicado a dados não separáveis linearmente, não encontra a solução desejada. Isso é evidenciado pela função objetivo (dual) que, aplicada a dados não linearmente separáveis, cresce arbitrariamente (BURGES, 1998). O principal problema desse classificador é que ele sempre constrói hipóteses que se baseiam na inexistência de erros de treinamento. Entretanto, para dados com ruídos, que geralmente implica em separação não linear, o mínimo para o risco esperado não pode ser calculado dessa forma, pois pode causar *overfitting*. Essas desvantagens motivaram o desenvolvimento de técnicas que permitem o tratamento de problemas não-linearmente separáveis via SVM.

3.4.3 Hiperplano ótimo para padrões não-separáveis

Considera-se agora o caso mais difícil, o de padrões não-separáveis. Dado um conjunto de dados de treinamento como este, não é possível construir um hiperplano de separação sem encontrar erros de classificação. Apesar disso, deseja-se encontrar um hiperplano ótimo que minimize a probabilidade de erro de classificação, calculada como a média sobre o conjunto de treinamento.

Diz-se que a margem de separação entre classes é suave se um ponto de dado (x_i, d_i) violar a seguinte condição (Equação 3.18):

$$d_i(w^T x_i + b) \geq 1, \quad i = 1, 2, \dots, N$$

Esta violação pode surgir de duas formas:

- O ponto de dado (x_i, d_i) se encontra dentro da região de separação, mas do lado correto da superfície de decisão, como ilustrado na figura abaixo 3.7(a).

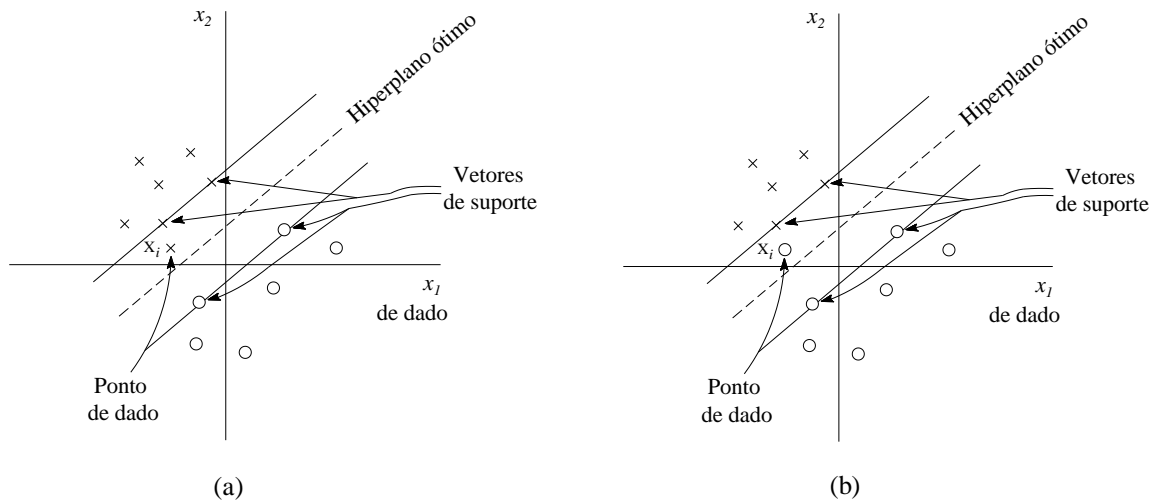


FIGURA 3.7(a) PONTO DADO DENTRO DA REGIÃO DE SEPARAÇÃO, MAS DO LADO CORRETO DA SUPERFÍCIE DE DECISÃO.

FIGURA 3.7(b) PONTO DADO NO LADO ERRADO DA SUPERFÍCIE DE DECISÃO.

FONTE: Haykin (1999)

- O ponto de dado (x_i, d_i) se encontra no lado errado da superfície de decisão, como ilustrado na figura acima 3.7(b).

Note a classificação correta no caso 1, mas uma classificação incorreta no caso 2.

Para preparar o terreno para um tratamento formal para o caso de pontos de dados não-separáveis, introduz-se um novo conjunto de variáveis escalares não-negativas, ξ_i , na definição do hiperplano de separação (i.e., superfície de decisão) como mostrado aqui:

$$d_i(w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \quad (3.27)$$

Essa estratégia permite tolerar ruídos e *outliers*¹⁰, considera mais pontos de treinamento além dos que estão na fronteira e permite a ocorrência de erros de classificação.

As ξ_i são chamadas de variáveis soltas ou de folga e medem o desvio de um ponto dado da condição ideal de separabilidade de padrões. Para $0 \leq \xi_i \leq 1$, o ponto de dado se encontra dentro da região de separabilidade, mas no lado correto da superfície de decisão, como ilustrado na figura 3.7(a). Para $\xi_i > 1$, ele se encontra do lado errado do hiperplano de separação, como ilustrado na figura 3.7(b). Os vetores de suporte são aqueles pontos de dados particulares que satisfazem a Equação 3.27 precisamente, mesmo se $\xi_i > 0$. Note que se um

¹⁰ Pontos muito distantes das classes a que pertencem.

exemplo com $\xi_i > 0$ for deixado de fora do conjunto de treinamento, a superfície de decisão não muda. Assim, os vetores de suporte são definidos exatamente do mesmo modo tanto para o caso linearmente separável como para o caso não-separável (HAYKIN, 1999).

Para tornar o problema matematicamente tratável, aproximamos o funcional $\Phi(\xi)$, escrevendo:

$$\Phi(\xi) = \sum_{i=1}^N \xi_i$$

Além disso, simplificamos a computação formulando o funcional a ser minimizado em relação ao vetor peso w , como segue:

$$\Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \quad (3.28)$$

A minimização do primeiro termo da Equação 3.28 está relacionada com a minimização da dimensão V-C¹¹ da máquina de vetor de suporte. Assim como para o segundo termo $\sum_i \xi_i$, ele é um limite superior para o número de erros de teste.

O parâmetro C controla o compromisso entre a complexidade da máquina e o número de pontos não-separáveis; por isso, pode ser visto como uma forma de parâmetro de “regularização”. Segundo Haykin (1999), o parâmetro C deve ser selecionado pelo usuário.

De qualquer forma, o funcional $\Phi(w, \xi)$ é otimizado em relação à w e ξ_i , sujeito à restrição descrita na Equação 3.27 e $\xi_i \geq 0$.

Pode-se agora formalizar o problema primal para o caso não-separável como:

1) Dada a amostra de treinamento $(x_i, d_i)_{i=1}^N$, encontre os valores ótimos de vetor peso w e do *bias* b de modo que satisfaçam à restrição:

$$\begin{aligned} d_i(w^T x_i + b) &\geq 1 - \xi_i \text{ para } i = 1, 2, \dots, N \\ \xi_i &\geq 0 \text{ para todo } i \end{aligned}$$

¹¹ A dimensão VC denominada em homenagem aos seus criadores Vapnik e Chervonenkis, pode ser definida como a medida da capacidade de aprendizado de uma classe de funções que classifica o maior número de amostras de treinamento (BURGES, 1998). Existem classes de funções com valores de dimensão VC (capacidades) diferentes, tais como: lineares, exponenciais, polinomiais.

2) É de modo que o vetor w e as variáveis soltas ξ_i minimizem o funcional de custo:

$$\Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i$$

onde C é um parâmetro positivo especificado pelo usuário.

Usando o método dos multiplicadores de Lagrange, pode-se formular o problema dual para padrões não-separáveis:

1) Dada a amostra de treinamento $(x_i, d_i)_{i=1}^N$, encontre os multiplicadores de Lagrange α_i que maximizam a função objetivo

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j$$

2) Sujeita às restrições:

$$\begin{aligned} \sum_{i=1}^N \alpha_i d_i &= 0 \\ 0 &\leq \alpha_i \leq C \end{aligned}$$

onde C é um parâmetro positivo especificado pelo usuário.

Note que nem as variáveis soltas ξ_i nem os multiplicadores de Lagrange aparecem no problema dual. O problema dual para o caso de padrões não-separáveis é dessa forma similar àquele para o caso simples de padrões linearmente separáveis, exceto por uma diferença pequena, mais importante. A função objetivo $Q(x)$ a ser maximizada é a mesma em ambos os casos. O caso não-separável difere do caso separável pelo fato de que a restrição $\alpha_i \geq 0$ é substituída pela restrição mais rigorosa $0 \leq \alpha_i \leq C$. Exceto por esta modificação, a otimização restrita para o caso não-separável e os cálculos dos valores ótimos do vetor peso w e do *bias* b procedem do mesmo modo como no caso linearmente separável.

A solução ótima para o vetor peso w é dada por

$$w_0 = \sum_{i=1}^{N_s} \alpha_{0,i} d_i x_i \quad (3.29)$$

onde N_s é o número de vetores suporte. A determinação dos valores ótimos do *bias* b , utilizam as condições de Kuhn-Tucker (KKT) definidas por:

$$\alpha_i [d_i(w^T x_i + b) - 1 + \xi_i] = 0 \quad i = 1, 2, \dots, N \quad (3.30)$$

e

$$\mu_i \xi_i = 0, 1, 2, \dots, N \quad (3.31)$$

Os μ_i são multiplicadores de Lagrange que foram introduzidos para forçar a não-negatividade das variáveis soltas ξ_i para todo i . No cálculo de ponto de mínimo (ponto de sela), a derivativa da função Lagrangeana com respeito as variáveis ξ_i é zero, sendo assim:

$$\alpha_i + \mu_i = C \quad (3.32)$$

Combinando as equações 3.30 e 3.31 nota-se

$$\xi_i = 0 \text{ se } \alpha_i = C \quad (3.33)$$

Como no caso separável, os pontos x_i para os quais $\alpha_i > 0$ são chamados de *vetores de suporte*. A diferença existente agora é que tem que distinguir entre os vetores de suporte para os quais $0 \leq \alpha_i \leq C$. Pode-se observar que as variáveis ξ_i não aparecem no problema dual. A função objetivo a ser minimizada é igual em ambos os casos.

Os multiplicadores de Lagrange são analisados da seguinte forma: quando $0 < \alpha_i \leq 1$, as amostras correspondentes constituem um vetor de suporte; se for $\alpha_i = 0$ tem-se uma amostra de treinamento que não participa da solução do problema, e caso seja $\alpha_i = C$ as amostras estão classificadas incorretamente (HAYKIN, 1999).

3.4.4 Funções *Kernel*

Com o conteúdo descrito nas seções anteriores, é possível, a partir deste ponto, descrever formalmente a construção de SVM para uma tarefa de Reconhecimento de Padrões. Segundo Santos (2002), SVM implementa basicamente duas operações matemáticas:

- 1) Mapeamento não-linear dos vetores de entrada x em um espaço de características Z com alta dimensão.
- 2) Construção de um Hiperplano de Margem Máxima no espaço de características.

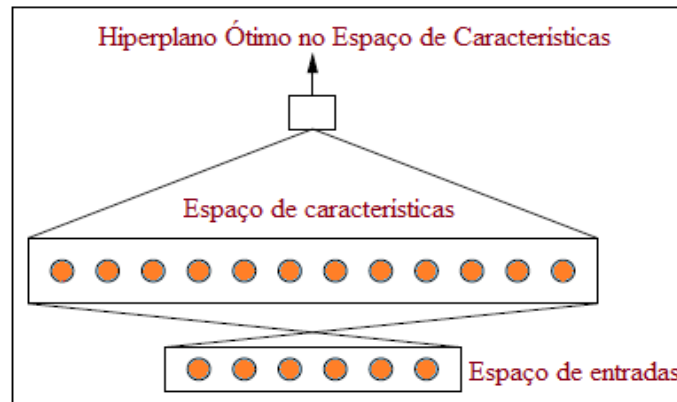


FIGURA 3.8: MAPEAMENTO DOS DADOS DO ESPAÇO DE ENTRADA NO ESPAÇO *FEATURE*
 FONTE: A autora, adaptado de Vapnik (1998)

A primeira operação é realizada de acordo com o teorema de Cover, que é descrito em (HAYKIN, 1999), da seguinte forma:

Um problema complexo de classificação de padrões tem mais probabilidade de ser separável linearmente em um espaço de alta dimensão do que em um espaço de baixa dimensão.

Pode-se fazer um mapeamento dos dados não-lineares em um espaço de dimensão maior, onde os dados tornam-se linearmente separáveis (CARVALHO, 2005).

Considerando-se um espaço de entrada em que os padrões não são linearmente separáveis, o teorema de Cover diz que esse espaço pode ser transformado em um novo espaço de características (*feature*), onde os padrões têm alta probabilidade de tornarem-se linearmente separáveis.

A segunda operação implementa a idéia de construir um Hiperplano de Margem Máxima de acordo com a teoria descrita na seção 3.4.3, mas com uma diferença fundamental: o Hiperplano de Margem Máxima agora é definido como uma função linear de vetores do espaço *feature*, e não do espaço de entrada original.

Foi observado por Vapnik (1998) que para a construção do Hiperplano de Margem Máxima no espaço *feature* Z não é necessário considerar tal espaço de forma explícita, e sim, apenas calcular os produtos internos entre os vetores de suporte e os vetores do espaço *feature*.

Para criar um espaço *feature*, por exemplo, um espaço polinomial de grau dois, seja x um vetor no espaço de entrada com n coordenadas, podendo ser escrito como $x = (x^1, x^2, x^3, \dots, x^n)$. O espaço *feature* teria suas coordenadas da seguinte maneira:

- O espaço de entradas: $z^1 = x^1, z^2 = x^2, \dots, z^n = x^n$ (n coordenadas);
- O espaço de entrada ao quadrado: $z^{n+1} = (x^1)^2, \dots, z^{2n} = (x^n)^2$ (n coordenadas);
- A multiplicação entre as dimensões do espaço de entrada $z^{2n+1} = x^1 x^2, \dots, z^N = x^n x^{n-1}$. $\left(\frac{n(n+1)}{2} \text{ coordenadas}\right)$

Ou seja, um espaço de dimensão n seria levado num espaço *feature* de dimensão $N = \frac{n(n+3)}{2}$. Se o problema tem dimensão pequena, é possível construir o espaço *feature*. No entanto, se o problema tiver dimensão maior, por exemplo, R^{200} , um polinomial de grau cinco geraria bilhões de coordenadas, tornando-se difícil tratá-lo computacionalmente.

Na função objetivo do problema dual do SVM existe um produto interno entre os pontos, o qual pode ser substituído por uma única função, denominada função *kernel* (ALES, 2008).

Um *kernel* é uma função K , tal que para todo $x, z \in X$

$$K(x, z) = \langle \phi(x), \phi(z) \rangle \quad (3.34)$$

onde ϕ é um mapeamento de X de um (produto interno) no espaço *feature* F (SHAWER-TAYLOR e CRISTIANINI, 2000).

Observe a Figura 3.9 abaixo:

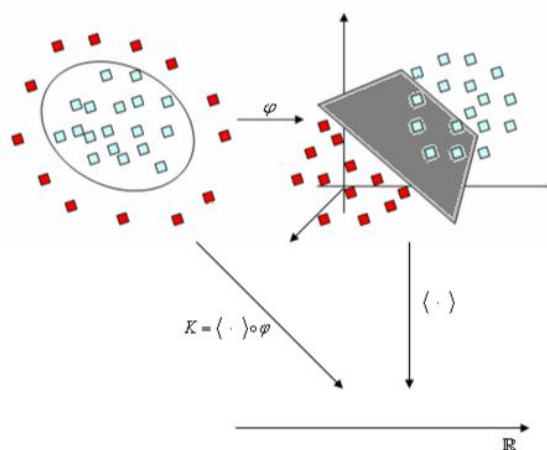


FIGURA 3.9: MAPEAMENTO REALIZADO PELA FUNÇÃO *KERNEL* EM UM ESPAÇO DE DIMENSÃO MAIOR
 FONTE: Carvalho (2005)

Os dados que estão no espaço de entrada não são separáveis por um hiperplano. A função φ faz uma transformação nos dados, de maneira que no espaço *features*, tornam-se separáveis.

Fazendo este produto interno sendo o usual, obtém-se uma superfície separadora linear, que separa apenas conjuntos separáveis. Isso depende apenas da função *kernel* escolhida.

A escolha da função *kernel* é de vital importância para o SVM. Segundo Shower-Taylor e Cristianini (2000) o problema de escolher uma arquitetura para uma aplicação de Rede Neural é substituído pela escolha do *kernel* adequado para o SVM. Portanto, devemos primeiro determinar quais as propriedades são necessárias para assegurar que ela é uma *kernel* para algumas características de espaço. Inicialmente, para uma função ser considerada *kernel* são necessárias as condições:

- 1) A função tem que ser simétrica, isto é:

$$K(x, z) = \langle \phi(x). \phi(z) \rangle = \langle \phi(z). \phi(x) \rangle = K(z, x) \quad (3.35)$$

- 2) Satisfazer as inequações de Cauchy-Schwarz:

$$\begin{aligned} K(x, z)^2 &= \langle \phi(x). \phi(z) \rangle^2 \leq \|\phi(x)\|^2 \|\phi(z)\|^2 \\ &= \langle \phi(x). \phi(x) \rangle \langle \phi(z). \phi(z) \rangle = K(x, x) K(z, z) \end{aligned} \quad (3.36)$$

Essas condições não são suficientes para garantir a existência do espaço *feature*. As funções *kernel* devem satisfazer uma terceira propriedade que é fornecida pelo Teorema de Mercer. A condição de Mercer para garantir que a função simétrica $K(x, z) = \sum_i \phi(x)_i \phi(z)_i$ seja um *kernel* é que a seguinte integral seja válida:

$$\iint K(x, z) g(x) g(z) dx dz > 0 \quad (3.37)$$

seja válida para todo $g \neq 0$ e

$$\int g^2(x) dx < \infty \quad (3.38)$$

A descrição deste teorema encontra-se em detalhes em Shower-Taylor e Cristianini (2000).

Segundo Nguyen *et al.*(2006), não existe uma função *kernel* que seja superior a qualquer outra função. O desempenho depende do tipo de dados de cada aplicação.

Dentre as funções *kernel* mais utilizadas destacam-se:

1) Produto interno usual:

$$K(x^i, x^j) = (x^i)^t \cdot x^j \quad (3.39)$$

2) Polinomial homogêneo:

$$K(x^i, x^j) = \left((x^i)^t \cdot x^j \right)^p \quad (3.40)$$

onde:

p é o grau do polinômio

3) Polinomial não-homogêneo:

$$K(x^i, x^j) = \left((x^i)^t \cdot x^j + k \right)^p \quad (3.41)$$

onde:

p é o grau do polinômio

k é uma constante especificada *a priori* pelo usuário

4) Sigmoidal:

$$K(x^i, x^j) = \tanh(\kappa x^i \cdot x^j + k) \quad (3.42)$$

onde:

κ é uma constante

k é uma constante negativa

Hsu *et al.*(2008) afirmam que este *kernel* não é válido para alguns parâmetros, em virtude de não ser um produto interno.

5) Gaussiano:

$$K(x^i, x^j) = e^{-\frac{\|x^i - x^j\|^2}{2\sigma^2}} \quad (3.43)$$

onde:

σ é um parâmetro especificado *a priori* pelo usuário.

Essa função também é conhecida como Função de Base Radial (*Radial Basis Function*), RBF.

A função *kernel* não altera muito o SVM, pelo menos não explicitamente. Com a introdução da função *kernel*, para encontrar os coeficientes α_i é necessário resolver agora o seguinte problema:

$$\text{Maximize:} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \quad (3.44)$$

$$\text{Sujeito a:} \quad 0 \leq \alpha_i \leq C \quad i = 1, \dots, N$$

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (3.45)$$

Essa funcional coincide com a funcional para encontrar o Hiperplano de Margem Máxima, exceto pela forma do produto interno, que anteriormente era representado unicamente por produto interno e agora passa a ser representado pela função *kernel* $K(x^i, x)$. Com a função *kernel*, a função de separação (ou de decisão) passa a ser representada como:

$$f(x) = \text{sinal} \left(\sum_{\text{vetores de suporte}} \alpha_i d_i K(x^i, x) + b \right) \quad (3.46)$$

Portanto, para reconhecer a qual conjunto um novo padrão pertence, é possível ser verificado apenas pelo sinal da função.

Segundo Shower-Taylor e Cristianini (2000), é esperado o menor número de vetores suporte para obter-se uma melhor generalização.

Para classificar os padrões do conjunto de treinamento como vetores suporte, verificam-se os valores dos respectivos Multiplicadores de Lagrange, sendo:

- Se $\alpha_i = 0, d^i \cdot f(x^i) > 1$ então x^i é considerado um vetor comum, que se situa do lado correto, na região de sua classe, e não influencia na construção do hiperplano separador ótimo.
- Se $0 < \alpha_i < C, d^i \cdot f(x^i) = 1$ então x^i é um vetor de suporte, situa-se sobre a margem da região de sua classe. É conhecido como vetor suporte *non-bound* (VS-NB).

Enfim, se $\alpha_i = C, d^i \cdot f(x^i) < 1$ então x^i é um vetor de suporte *bound* (VS-bound). Ele pode se localizar entre a margem e o hiperplano separador, caso $0 < \xi_i < 1$; pode estar na própria superfície de separação, caso $\xi_i = 1$; ou pode estar na região de classe oposta a sua, do outro lado da superfície de separação, caso $\xi_i > 1$ (SHOWER-TAYLOR e CRISTIANINI, 2000).

Nas figuras a seguir, observa-se a utilização de funções *kernel* para a separação entre classes:

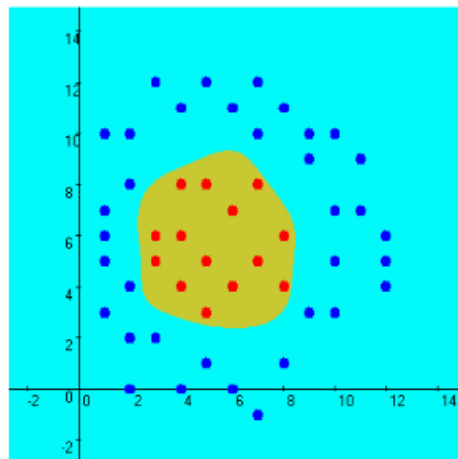


FIGURA 3.10: SEPARAÇÃO DOS DADOS
FONTE: Ales (2008)

Na Figura 3.10 acima, os dados de uma classe -1 (pontos em azul) estão formando uma envoltória ao redor dos dados da classe +1 (pontos em vermelho), sendo impossível encontrar um hiperplano separador (nesse caso, uma reta) que separasse os pontos corretamente. A utilização da função *kernel* (nesse caso, o *kernel* gaussiano) obteve a separação ótima dos 49 dados, generalizando a situação dada, com uma margem de segurança. Nesse treinamento, com parâmetros $\sigma=10$, obteve-se 18 vetores suportes (ALES, 2008).

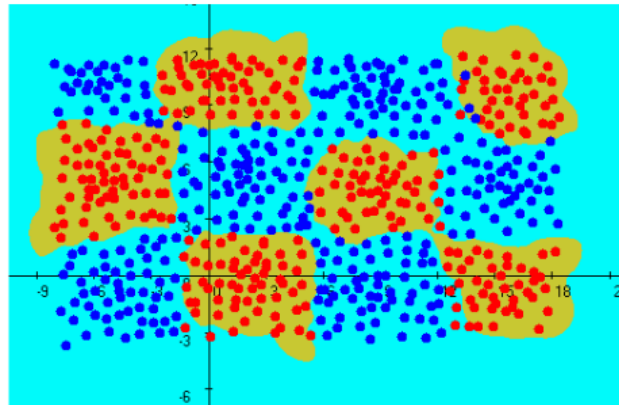


FIGURA 3.11: SEPARAÇÃO DOS DADOS EM SUB-REGIÕES
FONTE: Ales (2008)

Os dados apresentados na Figura 3.11 possuem características diferenciadas, com as classes possuindo sub-regiões, contendo 615 pontos. A utilização de uma função *kernel* consegue encontrar uma superfície de separação, tendo o menor número de erros de classificação.

Nesse exemplo, usou-se a função *kernel* gaussiano com parâmetros $C=1$ e $\sigma=1$, obtendo-se 4,59% de erro. Para esta separação, foram necessários 208 vetores de suporte, sendo 134 vetores suportes *bounds* (ALES, 2008).

3.4.5 Implementação ou Soluções do Modelo de Programação Quadrática do SVM

O treinamento do SVM pode ser reduzido à solução de um problema de otimização quadrático convexo sujeito às restrições lineares. Várias técnicas podem ser utilizadas para realizar a Programação Quadrática, dentre elas: Quase-Newton, Gradiente Conjugado, Método de Filtros, entre outras¹². Há vários pacotes computacionais para Programação Quadrática, gratuitos e comerciais.

As rotinas acima, de Programação Quadrática para pequenas bases de dados, são a melhor escolha, mas, para grandes bases de dados, há a necessidade de usar técnicas alternativas. Mas a estrutura do problema de otimização de SVM permite derivar algoritmos que possibilitam uma convergência de forma mais rápida e com menor necessidade de memória. São eles: *Chunking*, Decomposição, SMO (*Sequential Minimal Optimization*), entre outros¹³.

¹² Para maiores aprofundamentos teóricos consultar Friedlander, 1994.

¹³ Maiores detalhes em Osuna et al.(1997); Muller et al. (2001)

3.4.5.1 SMO (*Sequential Minimal Optimization*)

A modelagem do SVM envolve um problema quadrático, e para resolvê-lo existem alguns métodos clássicos como o Método do Gradiente, Método de Newton, Método Quase-Newton, Métodos de Filtro, entre outros (FRIEDLANDER, 1994). Esses métodos obtêm as soluções ótimas após um determinado número de iterações, mas existem alguns pacotes computacionais que agilizam a resolução do problema do SVM, como o MINOS, do *Stanford Optimization Laboratory*, o LOQO, o *MATLAB Optimization Toolbox*, o *SVMlight* de Joachims e o LIBSVM, de Chih-Chung Chang e Chih-Jen Lin. Neste trabalho, o algoritmo utilizado é o *Sequential Minimal Optimisation* (SMO) proposto por John Platt (PLATT, 1998), em virtude de contemplar o modelo original do SVM e por ser implementado em *Visual Basic* por Ales (2008), que será o programa utilizado para o treinamento e teste dos dados.

O melhor deste algoritmo (SMO) reside no fato de que o problema de otimização utiliza apenas duas variáveis a cada iteração e admite uma solução analítica, eliminando a necessidade de resolver um problema quadrático do SVM, dispensando um programa de otimização externo.

Essa solução é encontrada trabalhando com dois Multiplicadores de Lagrange de cada vez e mantendo os demais fixos. Considerando os dois multiplicadores a serem atualizados como α_1 e α_2 e suas respectivas classes d e d_2 , a condição inicial para esse algoritmo é que $\sum_{i=1}^l \alpha_i d_i = 0$, ou seja, quando um multiplicador for atualizado, essa condição obriga a ajustar o outro, para que a condição continue verdadeira.

A escolha dos dois pontos é determinado por uma heurística, enquanto a otimização dos dois multiplicadores é realizada analiticamente. Apesar da necessidade de fazer convergir mais iterações, cada iteração utiliza um número tão reduzido de operações que o algoritmo apresenta uma velocidade total de algumas ordens de magnitudes. Além da convergência do tempo, outra importante característica do algoritmo é que ele não precisa armazenar a matriz *kernel* na memória (SHAWER-TAYLOR e CRISTIANINI, 2000).

Para não violar a condição linear $\sum_{i=1}^l \alpha_i d_i = 0$, os novos multiplicadores de Lagrange devem respeitar:

$$\alpha_i d_i + \alpha_2 d_2 = constante = \alpha_1^{velho} d_1 + \alpha_2^{velho} d_2 \quad (3.47)$$

No espaço (α_1, α_2) , e definido por $0 \leq \alpha_1, \alpha_2 \leq C$.

O algoritmo primeiramente encontra o valor para α_2^{novo} e sucessivamente utiliza-o para obter α_1^{novo} .

Em virtude disso, os multiplicadores de Lagrange além de possuírem os valores limitados entre 0 e C, restrição do problema do SVM, estão limitados por duas constantes U e V, para manterem a condição inicial verdadeira.

Podendo ser restrito pela condição:

$$U \leq \alpha_2^{novo} \leq V$$

onde

$$U = \max(0, \alpha_2^{velho} - \alpha_1^{velho}) \quad (3.48)$$

$$V = \min(C, C - \alpha_1^{velho} + \alpha_2^{velho}),$$

se $d_1 \neq d_2$, e

$$U = \max(0, \alpha_1^{velho} + \alpha_2^{velho} - C), \quad (3.49)$$

$$V = \min(C, \alpha_1^{velho} + \alpha_2^{velho}),$$

se $d_1 = d_2$.

Onde α_i^{novo} é o novo valor do Multiplicador de Lagrange do ponto x^i e α_i^{velho} é o valor anterior. O valor da função em x^i que denota a função atual determinada pelos valores dos Multiplicadores de Lagrange e por b no estágio atual da aprendizagem é dado por:

$$f(x^i) = \sum_{j=1}^l \alpha_j d_j K(x^j, x^i) + b \quad (3.50)$$

O valor E_i determina a diferença de $f(x^i)$ e o padrão y_i a que pertence o ponto x^i , ou seja, é a distancia do ponto ao hiperplano atual dado pela atualização dos Multiplicadores de Lagrange, é dado por:

$$E_i = f(x^i) - d_i = [\sum_{j=1}^l \alpha_j d_j K(x^j, x^i) + b] - d_i \quad (3.51)$$

Este valor pode ser grande quando o ponto está classificado corretamente, por exemplo, se o padrão $d_1 = 1$ e $f(x_1) = 5$, então $E_1 = 4$

A quantidade adicional exigida é a segunda derivada da função objetivo ao longo da linha diagonal, que pode ser expressa por κ , definido por:

$$\kappa = K(x^1, x^1) + K(x^2, x^2) - 2K(x^1, x^2) = \|\phi(x^1) - \phi(x^2)\|^2 \quad (3.52)$$

onde

x^1 e x^2 são os pontos associados a α_1 e α_2 , respectivamente e $\phi(\cdot)$ é o mapeamento no espaço *feature*.

Sendo α_2^{aux} um valor truncado, ou seja, limitado por $U \leq \alpha_2^{novo} \leq V$, dado por:

$$\alpha_2^{aux} = \alpha_2^{velho} + \frac{d_2(E_1 - E_2)}{\kappa} \quad (3.53)$$

O máximo valor da função objetivo será obtido com o valor:

$$\alpha_2^{novo} = \begin{cases} V & \text{se } \alpha_2^{aux} > V, \\ \alpha_2^{aux} & \text{se } U \leq \alpha_2^{aux} \leq V, \\ U & \text{se } \alpha_2^{aux} < U, \end{cases} \quad (3.54)$$

onde

E_1 é dado na Equação 3.51, κ é dado pela Equação 3.52, e U e V são dados pela Equação 3.48 e 3.49. O valor de α_1^{novo} é obtido de α_2^{novo} como segue:

$$\alpha_1^{novo} = \alpha_1^{velho} + d_1 d_2 (\alpha_2^{velho} - \alpha_2^{novo}) \quad (3.55)$$

O SMO usa dois critérios para selecionar dois pontos ativos para garantir que a função objetivo aproveite um grande acréscimo, na atualização dos valores.

Têm-se duas heurísticas, uma para a escolha de α_1 e outra para escolha de α_2 . Na primeira heurística, o ponto x_2 é escolhido entre os pontos que violam as condições de KKT. O algoritmo atravessa por todo o conjunto de dados de treinamento que violam as condições de KKT e seleciona um para atualizar, isso é feito através de um dos testes $E_2 \cdot d_2 < -tol$ e $\alpha_2 < C$ ou $E_2 \cdot d_2 > tol$ e $\alpha_2 > 0$. Quando tal ponto é encontrado, utiliza-se a segunda

heurística para seleccionar o ponto x^1 , que deve ser escolhido de tal maneira que seja atualizado com x^2 , causando um grande acréscimo na função objetivo dual (ALES, 2008).

Para encontrar um bom ponto a heurística escolhe x^1 , maximizando o valor dado por $|E_1 - E_2|$, se E_2 é positivo, o SMO escolhe o x^1 com o menor E_1 , e se E_2 é negativo, então o SMO escolhe x^1 com o maior E_1 .

Se esta escolha falhar em obter um acréscimo significativo na função objetivo dual, o SMO experimenta cada ponto x^1 que tenha valores de α diferente dos limites, ou seja, $0 < \alpha < C$, começando aleatoriamente. Se ainda não houver progresso significativo, o SMO procura por todo o conjunto de dados de treinamento para encontrar um ponto x^1 adequado.

Se ocorrer alteração de valores, a heurística retorna para escolher outros pontos x^2 e x^1 , até que todos os pontos estejam obedecendo as condições de KKT. Caso contrário termina o processo (ALES, 2008).

A lista dos erros de todos os pontos de treinamento é mantida na memória para reduzir contas adicionais.

A solução satisfaz as condições de complementaridade de Karush-Kuhn-Tucker (KKT) que para classificar com a máxima margem entre os conjuntos deve obedecer

$$\alpha_l [d_l(\langle w, x^i \rangle + b) - 1] = 0 \text{ para } i = 1, 2, \dots, l,$$

ou seja,

$$\alpha_i [d_i f(x^i) - 1] = 0$$

Correspondente a:

$$\begin{aligned} \alpha_i = 0 & \Leftrightarrow d_i f(x^i) \geq 1 \\ 0 \leq \alpha_i \leq C & \Leftrightarrow d_i f(x^i) = 1 \\ \alpha_i = C & \Leftrightarrow d_i f(x^i) \leq 1 \end{aligned} \tag{3.56}$$

Para melhorar a velocidade de convergência, é possível uma escolha do conjunto dos dois pontos a serem otimizados, baseados na suas respectivas contribuições para progresso geral em direção à solução. Se a quantidade de contas exigidas para implementar a estratégia

de seleccionar o conjunto reduzir o número de iterações, então obtem-se uma taxa de convergência melhor.

Shawe-Taylor e Cristianini (2000) descreve três critérios de parada, que são detalhados a seguir: 1) Monitoramento o valor da função, especificamente o valor do crescimento a cada passo, ou seja, o treinamento para quando a taxa de crescimento da função objetivo for menor que certa tolerância, por exemplo, 10^{-9} . 2) Monitoramento das condições de KKT para o problema primal, a condição de KKT é necessária e suficiente para encontrar o ponto ótimo no problema quadrático. 3) Outra maneira para caracterizar a solução é por meio do *gap*¹⁴ entre as funções objetivas: dual e primal, porém, esse critério de parada é válido apenas quando se tem um hiperplano linear.

Resolvendo o SMO, o valor de b , do vetor w e dos erros E_i são calculados separadamente. Após cada iteração, em que as condições de KKT são satisfeitas para ambos os pontos x^1 e x^2 . Os valores podem ser atualizados analisando sempre o valor atual com o anterior da função $f(x^i) = \sum_{j=1}^l \alpha_j d_j K(x^j, x^i) + b$.

O valor de b para o ponto x^1 é definido por b_1 , o qual deve forçar a saída do SVM para d_1 quando a entrada for o ponto x^1 , dado por:

$$b_1 = -[E_1 + d_1(\alpha_1^{novo} - \alpha_1^{velho})K(x^1, x^1) + d_2(\alpha_2^{novo} - \alpha_2^{velho})K(x^1, x^2) + b^{velho}] \quad (3.57)$$

O valor de b para o ponto x^2 é definido por b_2 , o qual deve forçar a saída do SVM para d_2 quando a entrada for o ponto x^2 , dado por:

$$b_2 = -[E_2 + d_1(\alpha_1^{novo} - \alpha_1^{velho})K(x^1, x^2) + d_2(\alpha_2^{novo} - \alpha_2^{velho})K(x^2, x^2) + b^{velho}] \quad (3.58)$$

Se os valores b_1 e b_2 são iguais, então este será o novo valor de b , ou seja, $b^{novo} = b_1 = b_2$. Caso contrário, o intervalo entre b_1 e b_2 são todos os *thresholds*¹⁵ que são consistentes com as condições de KKT, portanto, o SMO escolhe o *threshold* que está no meio do intervalo, ou seja:

$$b^{novo} = \frac{b_1 + b_2}{2} \quad (3.59)$$

¹⁴ A diferença entre os valores da função objetivo do modelo primal e do modelo dual é conhecido como *gap* dual (SHAWER-TAYLOR e CRISTIANINI, 2000).

¹⁵ Um *threshold* ou limiar é um valor mínimo de alguma quantidade.

Quando os dados são separados linearmente, pode-se atualizar o valor do vetor w por:

$$w^{novo} = w^{velho} + d_1(\alpha_1^{novo} - \alpha_1^{velho})\vec{x}^1 + d_2(\alpha_2^{novo} - \alpha_2^{velho})\vec{x}^2 \quad (3.60)$$

Os erros E_i são atualizados a cada iteração por:

$$E_i^{novo} = E_i^{velho} + d_1(\alpha_1^{novo} - \alpha_1^{velho})K(x^1, x^i) + d_2(\alpha_2^{novo} - \alpha_2^{velho})K(x^2, x^i) + b^{novo} - b^{velho} \quad (3.61)$$

O objetivo da heurística SMO é obter os valores dos Multiplicadores de Lagrange para que estes tenham os erros tendendo a zero. A atualização da função objetivo pode ser feita pelo *gap*, que é a diferença entre a função objetivo atual e a anterior:

$$\begin{aligned} gap = & (\alpha_1^{novo} - \alpha_1^{velho}) + (\alpha_2^{novo} - \alpha_2^{velho}) - \left[\sum_{j=1}^2 \sum_{i=1}^l d_j d_i (\alpha_j^{novo} - \alpha_j^{velho}) K(x^j, x^i) \right]_{\substack{i \neq 1 \\ i \neq 2}} + \\ & - \frac{1}{2} d_1^2 (\alpha_1^{novo} - \alpha_1^{velho}) K(x^1, x^1) - \frac{1}{2} d_2^2 (\alpha_2^{novo} - \alpha_2^{velho}) K(x^2, x^2) + \\ & - d_1 d_2 [\alpha_1^{novo} \alpha_2^{novo} - \alpha_1^{velho} \alpha_2^{velho}] K(x^1, x^2) \end{aligned} \quad (3.62)$$

O SMO possui uma rotina de laços (*loops*), forçando a heurística a procurar por todos os pontos de treinamento os quais estão infringindo as condições de KKT.

Segue abaixo o fluxograma da heurística dividido em três partes principais. O primeiro fluxograma detalha a escolha do primeiro ponto, analisando se esse ponto está ou não violando as condições de KKT. No segundo fluxograma, tem-se a escolha do segundo ponto, que será otimizado com o primeiro ponto escolhido, no terceiro fluxograma, conforme as Figuras 3.12, 3.13 e 3.14 (ALES, 2008).

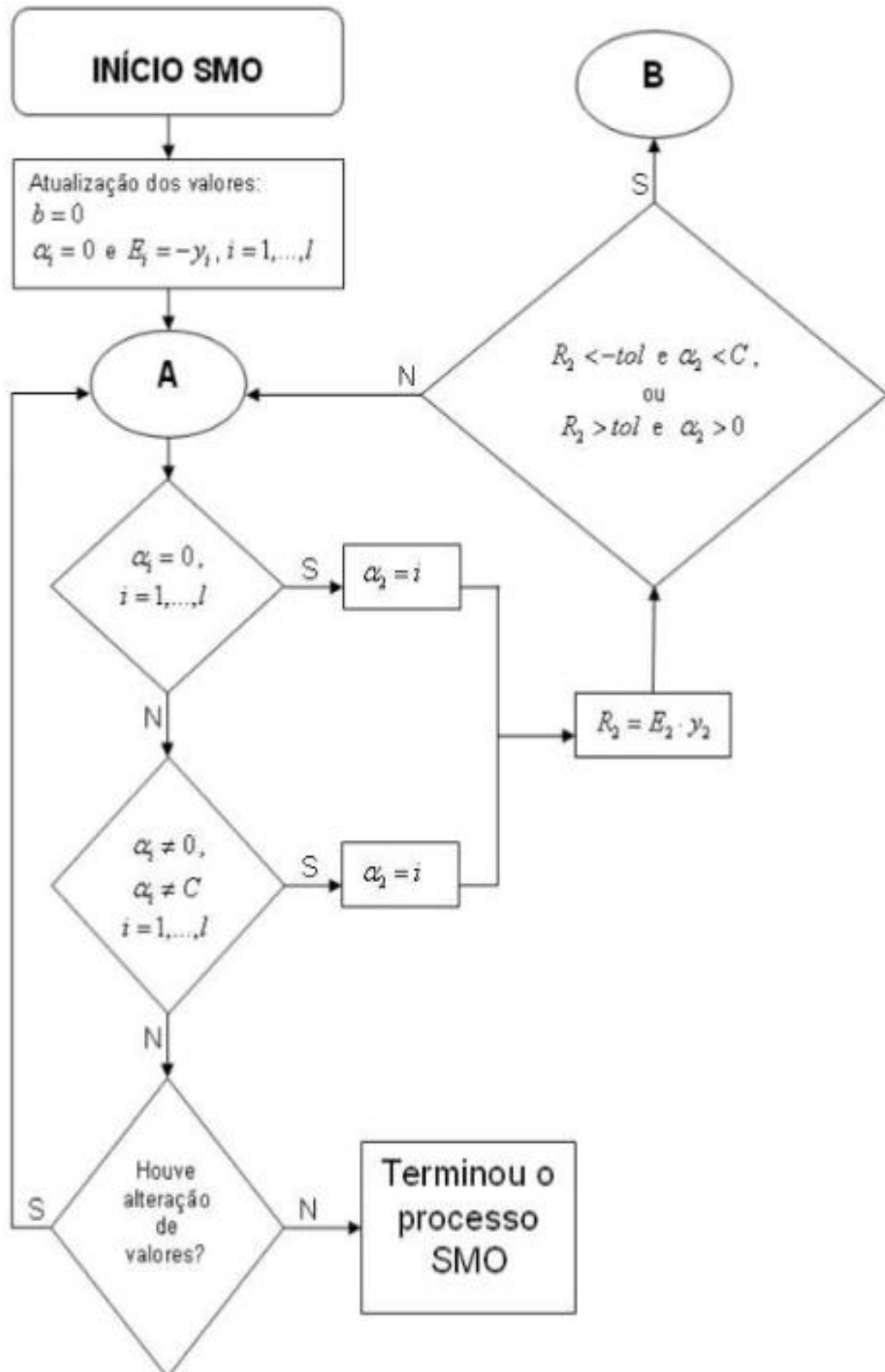


FIGURA 3.12: FLUXOGRAMA DA ESCOLHA DO PRIMEIRO PONTO
 FONTE: Ales (2008)

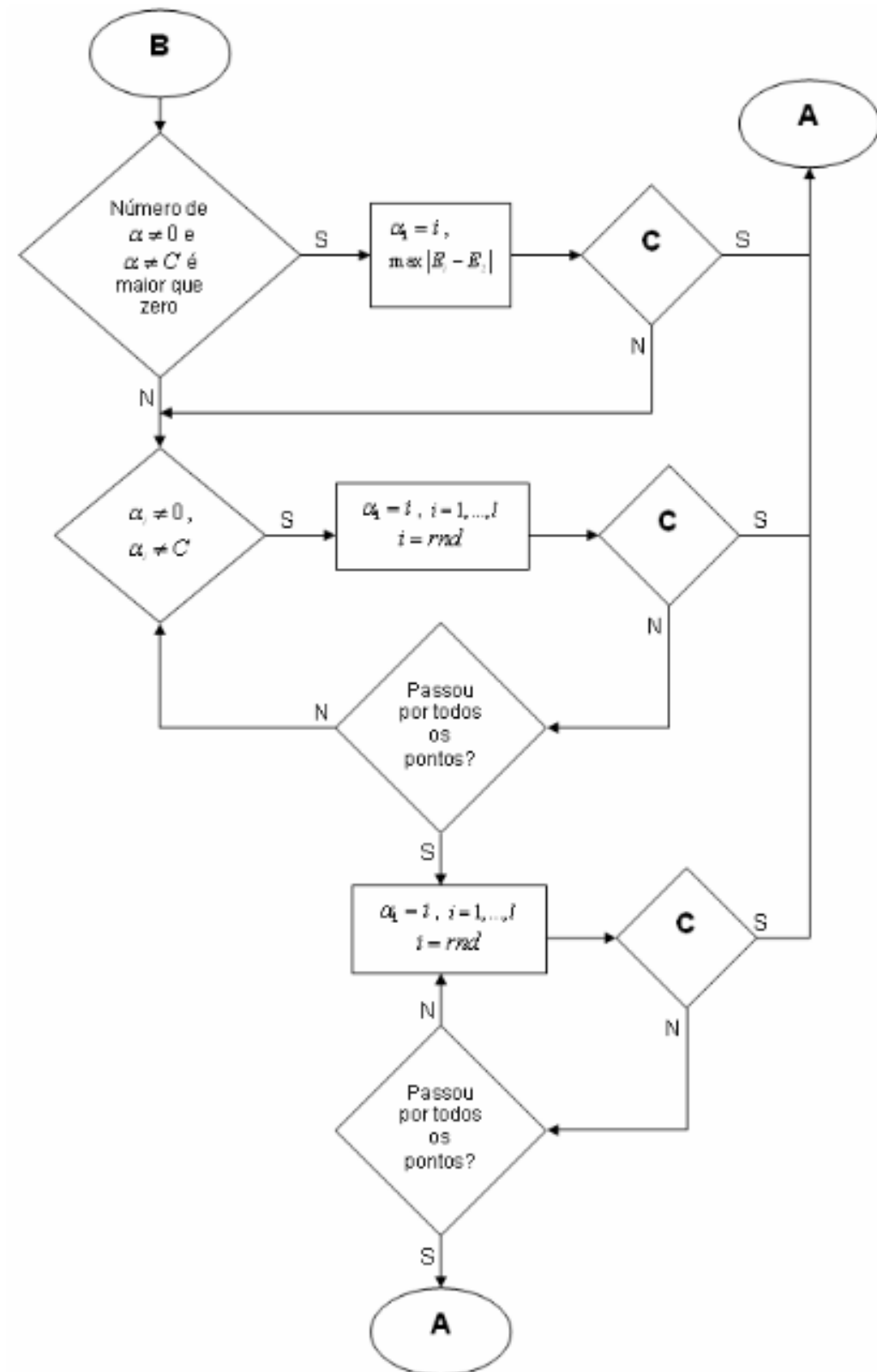


FIGURA 3.13: FLUXOGRAMA DA ESCOLHA DO SEGUNDO PONTO
 FONTE: Ales (2008)

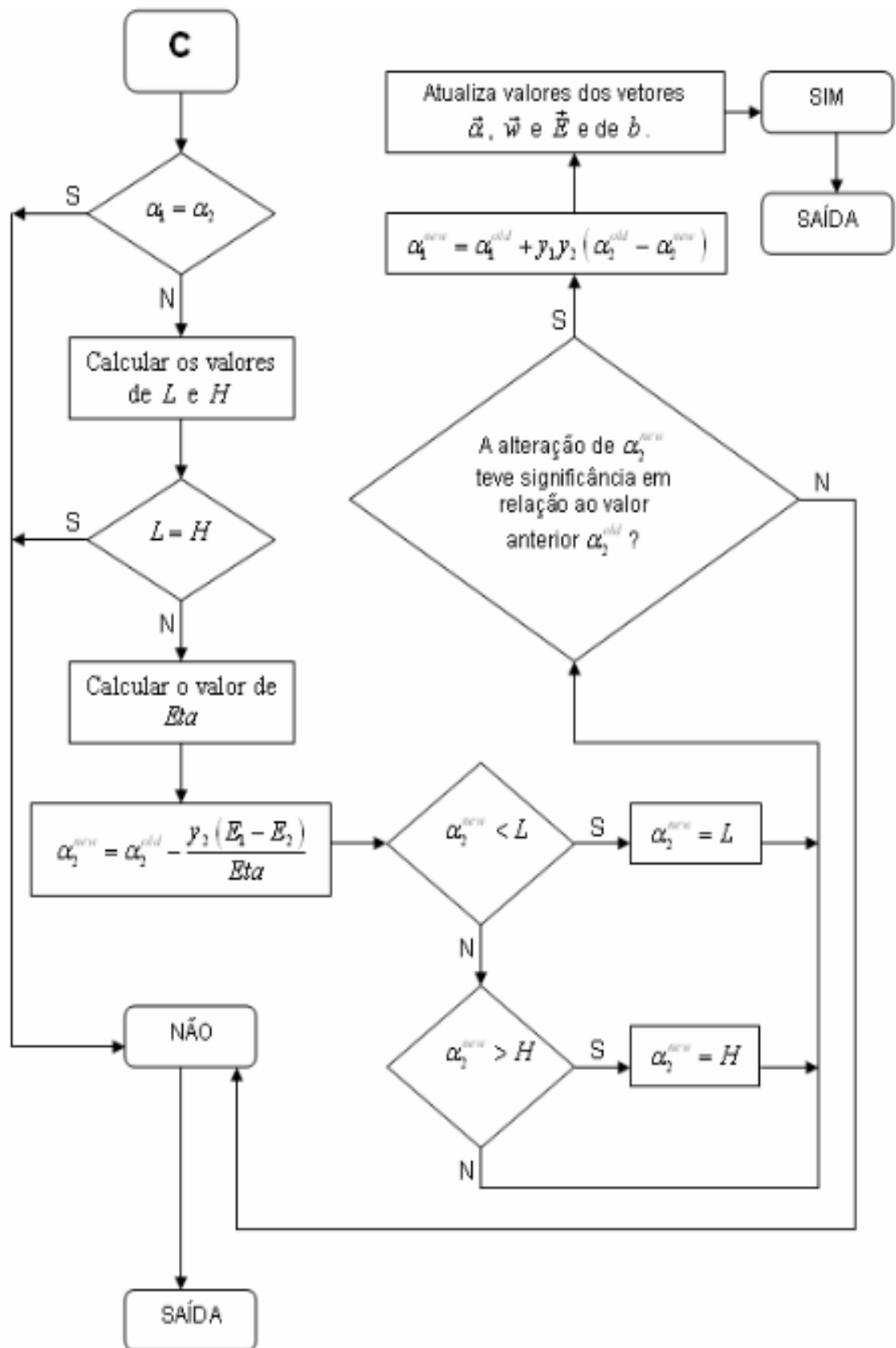


FIGURA 3.14: FLUXOGRAMA DA OTIMIZAÇÃO DOS PONTOS ESCOLHIDOS
 FONTE: Ales (2008)

CAPÍTULO IV

4 METODOLOGIA E RESULTADOS

Neste Capítulo, as técnicas do Capítulo III serão utilizadas na resolução do problema apresentado. Para que as técnicas sejam efetivamente utilizadas, é fundamental que se faça alguns comentários a respeito da origem das informações (banco de dados) e as adequações realizadas nesses dados para que pudessem ser utilizados nos algoritmos escolhidos.

O treinamento e o teste foram divididos em dois grandes grupos: no primeiro treinamento, foram usadas todas as variáveis do problema (normalizadas), conforme questionário; no segundo momento, o treinamento foi realizado após uma Análise Fatorial, com extração dos componentes principais, reduzindo o problema apenas aos fatores que explicam a maior parte dos dados.

4.1 DELIMITAÇÃO DO PROBLEMA

No desenvolvimento desta pesquisa, foram utilizados dados reais sobre os clientes de uma agência bancária na cidade de Wenceslau Braz, no Estado do Paraná, com unidades em todo o Brasil. Esta agência detêm uma grande fatia do mercado de pessoas jurídicas (empresas) da cidade, no que diz respeito ao crédito bancário.

A agência bancária oferece aos seus clientes linhas de crédito tanto para capital de giro quanto para investimentos. A clientela da carteira é constituída de micro, pequenas e médias empresas. As grandes empresas não fazem parte da carteira da agência em questão, pois este banco possui agência especializada para este atendimento.

É importante salientar, ainda, que as informações aqui apresentadas foram parcialmente transformadas, de forma a preservar o sigilo dos dados, mas sem alterar o seu significado.

4.2 SELEÇÃO DA AMOSTRA

Foi utilizada uma amostra de 199 empresas, das quais 64 são inadimplentes e 135 são adimplentes. Desta amostra foram extraídas as 25 informações, escolhidas pela necessidade do preenchimento do cadastro requerido pela agência bancária, de acordo com o Apêndice A deste trabalho.

As 25 informações que constituirão as variáveis do problema que esta sendo analisado são apresentadas a seguir.

- A: existência de restrições em nome da empresa

1 = SIM 2 = NÃO

- B: Existência de restrições baixadas nos últimos cinco anos em nome da empresa

1 = SIM 2 = NÃO

- C: Tempo de conta no Banco

Valor Numérico em Meses

- D: Setor de Atividade

1 = COMÉRCIO 2 = INDÚSTRIA 3 = SERVIÇOS

- E: Tempo de Atividade

1 = MAIS DE 9 ANOS 2 = DE 6 A 9 ANOS 3 = DE 3 A 5 ANOS

4 = DE 1 A 2 ANOS 5 = MENOS DE 1 ANO

- F: Número de funcionários

Valor Numérico

- G: Sede de empresa (imóvel)

1 = PRÓPRIO 2 = ALUGADO 3 = CEDIDO

- H: Nome do Bairro

1 = CENTRO

2 = OUTROS

3 = CEDIDO

- I: Principais clientes

1 = PESSOAS FÍSICAS

2 = PESSOAS JURÍDICAS

3 = MISTO

- J: Faturamento bruto anual

Valor Numérico

- K: Cliente em outro banco

1 = SIM

2 = NÃO

- L: Bens imóveis

Valor Numérico

- M: Bens móveis

Valor Numérico

- N: Seguro Empresarial

1 = SIM

2 = NÃO

- O: Aplicações financeiras no banco

1 = SIM > 8000

2 = SIM 4000 A 8000

3 = SIM 2000 A 4000

4 = SIM < 2000

5 = NÃO

- P: Vendas a prazo

1 = MENOS DE 20%

2 = MAIS DE 20%

- Q: Experiência de crédito no Banco

1 = SIM > 2 ANOS

2 = SIM < 2ANOS

3=NÃO

- R: Histórico da conta corrente

1 = NORMAL

2 = CHEQUES DEVOLVIDOS

3 = CLIENTE NOVO

4 = PEQUENOS ATRASOS FREQUENTES

- S: Sócios da empresa possuem restrições

1 = SIM

2 = NÃO

- T: Sócios da empresa tiveram restrições baixadas nos últimos cinco anos

1 = SIM

2 = NÃO

- U: Sociedade entre cônjuges

1 = SIM

2 = NÃO

- V: Existência de bens imóveis em nome dos sócios

Valor Numérico

- W: Existência de bens móveis em nome dos sócios

Valor Numérico

- X: Risco atribuído pelo Banco

1 = A

2 = B

3 = C

4 = D

5 = E

- Y: Resultado

1 = ADIMPLENTE

2 = INADIMPLENTE

Antes de analisar as informações do banco de dados, deverão ser escolhidas as variáveis que entrarão na análise, podendo vir a integrar o modelo final.

As variáveis V e W foram descartadas da base de dados, uma vez que as respostas estavam incompletas. A variável X também foi descartada, pois, é a classificação que o banco faz do cliente, e poderia influenciar no treinamento. A variável Y é a resposta do problema, classificada como variável dependente.

Após esta pré-análise, restaram 21 variáveis independentes e uma dependente(Y). A variável dependente foi transformada: código 1 para adimplente e 0 para inadimplente.

Concluídas as modificações, a nova matriz foi lançada em uma planilha eletrônica e ficou com o seguinte formato: 199 x 21, ou seja, 199 empresas com 21 componentes (características).

QUADRO 4.1: RESPOSTAS DO QUESTIONÁRIO APLICADO PELO BANCO ÀS EMPRESAS

n\v	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	2	2	615	1	1	30	1	1	1	6803	1	0	31	1	1	1	3	1	2	2	2
2	2	2	280	1	1	78	1	1	2	35002	1	0	0	1	1	2	3	1	2	2	2
3	2	2	216	1	1	8	1	1	1	1824	1	0	0	1	5	2	1	1	2	1	2
4	2	1	96	1	1	8	1	1	1	2055	1	0	0	1	5	2	1	1	2	1	2
5	2	2	108	1	1	2	1	1	1	629	1	0	0	1	2	2	1	1	2	2	2
6	2	1	120	1	2	2	1	1	1	121	1	0	0	2	5	1	3	1	2	2	2
7	2	2	216	1	1	5	1	1	1	3701	1	0	0	1	1	2	1	1	2	2	2
8	2	2	127	1	1	1	1	1	1	415	1	0	0	1	1	1	1	1	2	2	2
9	2	1	57	1	3	2	2	1	1	441	2	0	0	2	5	2	1	2	2	1	2
10	2	1	88	1	2	1	1	1	1	425	1	0	0	2	5	2	1	2	2	1	2
11	2	2	184	1	1	12	1	1	1	2018	1	0	0	1	1	2	1	1	2	2	2
.
.
198	1	1	12	1	3	4	2	2	1	78	1	0	0	2	5	1	1	2	1	1	2
199	1	1	12	1	2	2	1	1	1	90	1	0	0	2	5	2	3	2	2	2	2

FONTE: A autora (2009)

4.3 PADRONIZAÇÃO OU NORMALIZAÇÃO DOS DADOS

O processo de normalização¹⁶ deve ser feito quando os elementos de uma matriz de padrões estão com grandezas diferentes, o que pode tornar uma matriz mal condicionada¹⁷, possibilitando erros (BURDEN e FAIRES, 2003).

Uma importante vantagem desse processo é evitar números com escalas grandes dominando os números em escala menor, facilitando os cálculos durante o processo de Regressão Logística, Redes Neurais e *Support Vector Machine*.

¹⁶ Normalização é conhecida pelo termo *Scaling*.

¹⁷ Uma matriz é mal condicionada se mudanças relativamente pequenas em seus elementos podem causar mudanças relativamente grandes na solução.

A normalização é a substituição de valores absolutos por valores relativos, ou seja, transformação de valores de um atributo para uma escala reduzida.

A normalização escolhida foi a *Z-score*, que utiliza a média (A') e o desvio padrão (σ_A) dos valores do atributo para realizar a normalização, conforme equação 3.45, utilizando para isto a função *zscore* do *software MATLAB*.

$$v' = \frac{v - A'}{\sigma_A} \quad (3.45)$$

4.4 TREINAMENTO E TESTE

Para a obtenção de um bom resultado, devem ser feitas várias partições diferentes no mesmo conjunto de dados em conjuntos de treinamento e de teste. Este método, chamado de *cross validation*, fornece uma estimativa mais realista do erro a ser cometido pelo método de classificação utilizado. Deve-se manter nesta divisão a mesma proporção entre as classes para os conjuntos de treinamento e validação (MARTINELLI, 1999).

Conforme recomenda Haykin (1999) o número de partições diferentes no mesmo conjunto é 10, escolhendo-os de forma aleatória para a separação das amostras de treinamento e teste. Então, optou-se em utilizar o mesmo número de partições proposto pelo autor em um único conjunto de dados.

Em relação à divisão entre treinamento e teste, não existe nenhuma regra rígida estabelecida, e alguns autores preferem partições na proporção 60%-40% ou 75%-25% entre os grupos de treinamento e validação, respectivamente (HAIR *et al.*, 2005). Pesquisando trabalhos anteriores, observou-se diversas divisões: Cunico (2005) optou em dividir os conjuntos na proporção 70%-30%, Lemos (2003) na proporção 90%-10%, ambos obtendo resultados satisfatórios.

Neste trabalho, optou-se em separar os dados na proporção 80-20 para o conjunto de treinamento e teste, respectivamente, para observar se os resultados também seriam satisfatórios.

A amostra de teste, reservada para posterior validação dos modelos construídos, ficou composta por 27 clientes adimplentes e 13 clientes inadimplentes. A amostra de treinamento ficou formada, portanto, por 108 clientes adimplentes e 51 clientes inadimplentes.

Cada amostra ficou armazenada em uma planilha eletrônica. Separadas as matrizes de treinamento e teste, as técnicas escolhidas foram empregadas e os resultados seguem nos próximos itens.

4.5 TÉCNICA DE REGRESSÃO LOGÍSTICA

Para a implementação da técnica de Regressão Logística, utilizou-se o *software Statistica*, pois, além de apresentar recursos de cálculos e gráficos, é de fácil manuseio e utiliza pouco tempo de processamento.

A análise dos dados foi feita alocando no *software* a matriz de treinamento acrescentada da coluna das respostas. Portanto, a matriz a ser analisada ficou com formato 108 x 21.

O *software* possui um pacote especial na opção *Nonlinear Estimation*, que depois de alimentado com a matriz oferece a possibilidade de várias análises estatísticas, e entre elas está a Regressão Logística (*Logistic Regression*). Para a utilização do *software*, deverão ser informadas as variáveis independentes, dependente e o método utilizado para estimativa dos parâmetros.

4.5.1 Estimativa dos Parâmetros

Os parâmetros $\beta_0, \beta_1, \beta_2, \dots, \beta_{21}$ foram estimados pelo próprio pacote *Statistica*, com a utilização do método da máxima verossimilhança, num contexto de mínimos quadrados não-lineares, aplicando-se o Algoritmo de Quase-Newton.

Os resultados dos parâmetros seguem no quadro 4.2.

QUADRO 4.2: BETAS ESTIMADOS PELO MÉTODO DA MÁXIMA VEROSSIMILHANÇA

Modelo de Regressão Logística										
Parâmetros estimados										
β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
3,466	10,870	2,735	-2,062	0,004	-0,373	1,773	1,302	-1,751	1,039	-0,082
β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}	β_{17}	β_{18}	β_{19}	β_{20}	β_{21}
-1,967	0,1	1,983	0,166	-0,263	-0,178	-0,703	-0,333	6,485	0,965	1,852

FONTE: A autora (2009)

4.5.2 Classificações Corretas

A matriz de confusão contendo os acertos e os erros na fase de treinamento é gerada no próprio pacote *Statistica* e segue um exemplo na figura 4.1. A quantidade de empresas classificadas como 0 (inadimplentes) e que são 0 (inadimplentes) é de 51 enquanto que as quantidades de empresas classificadas como 1(adimplentes) e que são 0 (inadimplentes) é de 0. Segue o mesmo raciocínio para as empresas adimplentes.

	Pred. 0	Pred. 1	Percent Correct
Observed 0	51	0	100.0000
Observed 1	0	108	100.0000

FIGURA 4.1: MATRIZ DE CONFUSÃO GERADA PELO SOFTWARE STATISTICA
FONTE: A autora (2009)

4.5.3 Equações da Logit Estimada e do Modelo Estimado

De posse dos parâmetros betas estimados na análise de regressão, foi possível confeccionar as equações do modelo estimado e da logit estimada e que podem agora ser utilizadas para realizar a fase de teste. As equações seguem abaixo.

Logit estimada:

$$\hat{g}(x) = 3,4660 + 10,8708x_1 + 2,7351x_2 - 2,0628x_3 + \dots + 6,4854x_{19} + 0,9656x_{20} + 1,8529x_{21}$$

Modelo Estimado

$$\hat{\pi}(x) = \frac{e^{3,4660+10,8708x_1+2,7351x_2-2,0628x_3+\dots+6,4854x_{19}+0,9656x_{20}+1,8529x_{21}}}{1 + e^{3,4660+10,8708x_1+2,7351x_2-2,0628x_3+\dots+6,4854x_{19}+0,9656x_{20}+1,8529x_{21}}}$$

Cada linha da matriz de teste com suas respectivas variáveis é substituída em $\hat{\pi}(x)$, possibilitando o cálculo da variável binária que indicará se o indivíduo em estudo pertence a população “0” ou “1”.

4.5.4 Resultados da Regressão Logística

Depois de estimados os parâmetros, é necessário testar o modelo. Durante esta fase, o conjunto de teste é utilizado para verificar a performance com dados que não foram utilizados no treinamento. A performance do modelo, calculada nesta fase, é uma boa indicação de sua performance real.

O teste foi realizado utilizando uma planilha eletrônica, e o quadro 4.3 indica os resultados parciais da matriz de teste utilizando os parâmetros estimados na regressão e a função logística. Apresenta também o resultado da função logit e do modelo estimados.

QUADRO 4.3: RESULTADOS PARCIAIS DA TÉCNICA DE REGRESSÃO LOGÍSTICA

Núm.	Modelo Estimado	Logit	Realidade	Resultado
1	1,000271056	10,26464364	1	correto
2	1,000708452	23,76617127	1	correto
3	1,000512232	17,18648905	1	correto
4	1,00040489	13,62535411	1	correto
5	1,000364216	12,3637014	1	correto
6	1,000392004	13,21379014	1	correto
7	1,000517223	17,35374313	1	correto
8	1,000504833	16,93865217	1	correto
9	1,000270242	10,2521065	1	correto
10	1,000439957	14,77394598	1	correto
.
.
14	1,000643861	21,60009525	1	correto
15	1,0007328	24,58268333	1	correto
16	1,000522587	17,53349668	1	correto
17	1,000460391	15,45301151	1	correto
18	2,24237E-13	-29,12608077	1	errado
.
.
37	2,8039E-12	-26,60001634	0	correto
38	2,06237E-11	-24,60458607	0	correto
39	5,31015E-11	-23,65882295	0	correto
40	0,000552223	-7,501008262	0	correto

FONTE: A autora (2009)

Os resultados finais da técnica de regressão logística podem ser visualizados no quadro 4.4.

QUADRO 4.4: RESULTADO FINAL DA REGRESSÃO LOGÍSTICA

Regressão Logística					
Média acerto treinamento	Desvio padrão	Média acerto teste	Desvio padrão	Média de acerto do modelo	Tempo computacional (s)
100,00%	0,00%	96,99%	3,02%	98,50%	menor que 0,1

FONTE: A autora (2009)

Observa-se no quadro 4.4 que o modelo teve desempenho de 100% na fase de treinamento e desvio padrão de 0,00%, e na fase de teste teve 96,99% de desempenho e desvio padrão de 3,02%. Logo, a média de acerto do modelo é de 98,50%.

O tempo computacional mostrou-se satisfatório pois, nos 10 conjuntos treinados, o tempo foi menor que 0,1 segundos.

4.6 TÉCNICA REDES NEURAIIS

Para a implementação da técnica de Redes Neurais, foi utilizado o *software MATLAB – Neural Networks Toolbox*.

Existem duas formas de utilização das Redes Neurais no *MATLAB*: linha de comando *m-files* e interface gráfica (*NNTool*). Optou-se pela interface gráfica, por ser de fácil manuseio e visualização dos processos da rede neural.

A alimentação dos dados foi feita através do comando *Import Data*, que importa os dados diretamente da planilha eletrônica. Portanto, foram alocados os dados em três matrizes: a matriz A, contendo as entradas dos dados do treinamento (159 X 21), a matriz B, contendo as respostas dos dados de treinamento (159 X 1), e a matriz C, contendo as entradas dos dados de teste (40 X 21). Através dos comandos do *MATLAB*, todas as matrizes foram transpostas.

4.6.1 Modelo Implementado

Optou-se por uma rede *backpropagation* com correção de erro pelo algoritmo Levenberg-Marquadt¹⁸, que faz parte do *default* da função utilizada no *MATLAB*. Também foi necessário informar ao *software* o número de neurônios na camada de entrada, na camada escondida e na saída, bem como a função de ativação, o número de iterações e a margem de erro desejada.

A quantidade de neurônios da camada de entrada é igual ao número de variáveis utilizadas, que no caso deste trabalho é igual a 21, a quantidade de neurônios na camada de saída é igual a 1, e a função de ativação utilizada foi a sigmoideal (*logsig*).

¹⁸ Levenberg- Marquardt: é um algoritmo de aproximação que utiliza o método de Gauss-Newton modificado pela introdução de um parâmetro μ que funciona como fator de estabilização no treinamento da rede.

A técnica de Redes Neurais foi executada várias vezes, sempre variando a topologia da rede, ou seja, o número de neurônios na camada escondida ou ainda o número de iterações.

Após o treinamento da rede, é possível realizar uma simulação no próprio software com o conjunto de teste e obter as respostas para este conjunto.

Para verificar os resultados dos conjuntos de Teste das Redes Neurais, foi elaborada uma nova planilha eletrônica.

4.6.2 Resultados Obtidos Redes Neurais

O quadro 4.5 demonstra os resultados obtidos com a técnica de Redes Neurais:

QUADRO 4.5: PERFORMANCE DAS REDES TREINADAS

Rede	Média acerto treinamento	Desvio padrão	Média acerto teste	Desvio padrão	Número de neurônios	Número de iterações	Função	Tempo computacional (s)
1	91,37%	5,60%	87,78%	7,78%	6	30	logit	menor que 0,1
2	86,53%	12,75%	82,45%	14,90%	6	50	logit	menor que 0,1
3	91,42%	5,67%	89,09%	6,08%	6	100	logit	menor que 0,1
4	87,49%	13,92%	79,89%	17,91%	8	30	logit	menor que 0,1
5	93,39%	7,97%	90,16%	10,55%	8	50	logit	menor que 0,1
6	90,80%	7,85%	84,33%	11,56%	8	100	logit	menor que 0,1
7	91,08%	4,56%	84,69%	10,40%	10	100	logit	menor que 0,1
8	86,36%	10,16%	78,56%	14,65%	12	100	logit	menor que 0,1

FONTE: A autora (2009)

No quadro 4.5, pode-se destacar a rede 5, que possui acerto de 93,39% e 90,16% e desvio padrão de 7,99% e 10,55%, nas fases de treinamento e teste, respectivamente.

O tempo de execução das redes também pode ser destacado, pois as redes treinadas obtiveram tempos menores do que 0,1 segundos, o que significa baixo tempo computacional.

Sendo assim, a rede eleita é a número 5, por ter o melhor desempenho dentre as redes treinadas, baixo tempo computacional e média de acerto do modelo de 91,78%.

4.6.3 Pesos e bias

Os pesos e bias da rede são importantíssimos, pois são os coeficientes de multiplicações e adições que serão utilizados juntamente com as funções logit, para predizer se um novo indivíduo, com suas características individuais, pertence ao grupo dos adimplentes ou inadimplentes.

Desta forma, seguem no quadro 4.6, a título de ilustração, os resultados alcançados com a rede 5 no primeiro conjunto de treinamento.

QUADRO 4.6: PESOS E BIAS DA REDE Nº 5

Pesos da camada intermediária: 8 neurônios							
w1 = matriz 8 x 21							
-3,0287	1,5796	0,13371	-0,30823	2,0096	0,89479	-1,1575	
0,43394	-0,83417	0,76566	0,98236	2,726	-4,2674	-0,78561	
-6,7556	-0,84403	-0,22711	-0,252	-1,1229	0,7095	0,074214	
-4,6344	1,2925	-2,2014	-1,6878	-1,3141	-0,80476	0,7053	
-0,6625	1,5302	0,90828	0,36839	0,81397	1,1177	-1,0962	
0,58831	-0,071553	-0,044476	-4,7819	3,2269	-0,61163	5,1379	
4,7332	1,3952	2,524	1,7634	2,5519	0,22397	0,402	
-0,21258	3,089	2,7117	-1,1114	1,8886	0,54914	-1,6544	
-0,42332	0,19719	-1,8915	0,0040709	0	-0,27146	0,6898	
0,13758	-1,0755	1,024	2,5377	0	-0,28146	-0,91593	
1,1977	-0,61591	0,53556	0,38429	0	0,28337	-0,99149	
-5,5635	0,96542	-1,1667	-1,3844	0	-0,13525	0,83631	
-3,5635	-0,16637	-0,41587	-1,0755	0	0,096157	1,3939	
2,3687	-0,48422	-1,8003	-0,1907	0	0,49521	-1,8407	
-0,3427	0,42338	-1,4416	-0,79712	0	-0,1095	-0,035409	
-7,7788	2,048	0,043921	-3,5439	0	0,33909	0,78688	
0,077284	0,71392	-0,43993	0,96831	-0,80713	-2,0173	1,7748	
3,0824	2,3619	2,1618	-1,798	-3,6738	0,62548	-0,10981	
1,1775	-1,858	0,63476	0,31478	-5,1511	0,10462	-1,6724	
0,32444	-0,74891	-4,4081	-0,83908	-3,3326	0,67452	2,2321	
-1,0762	-1,5253	-1,2567	1,3927	2,0121	-0,036933	-2,4053	
-0,90831	-2,0102	-2,8041	1,079	-5,0563	5,0364	0,38266	
-1,8933	2,0645	-4,4036	-3,1069	1,4147	-0,74854	-0,36041	
0,95551	-2,1068	3,1612	-0,59291	-2,6681	-1,084	-1,5055	
Pesos da camada de saída: 1 neurônio							
w2 = vetor 1 x 8							
-1,0781	0,034154	-2,065	-0,086509	0,043665	0,026919	0,16283	0,045679
Bias da camada intermediária e da saída							
b1 = vetor 8 x 1	b2	3,0965					
6,644							
1,8352							
-0,59837							
2,7628							
0,042945							
1,3405							
-2,1197							
-0,019766							

FONTE: A autora (2009)

4.7 TÉCNICA *SUPPORT VECTOR MACHINES* (SVM's)

Para a implementação do *Support Vector Machine*, foi utilizado um programa desenvolvido por Ales (2008) em linguagem *Visual Basic*¹⁹ com o uso do algoritmo SMO, possibilitando a verificação de todo o processo e das superfícies de separação quando os dados estiverem no plano cartesiano.

A interface do programa é mostrada na Figura 4.2.

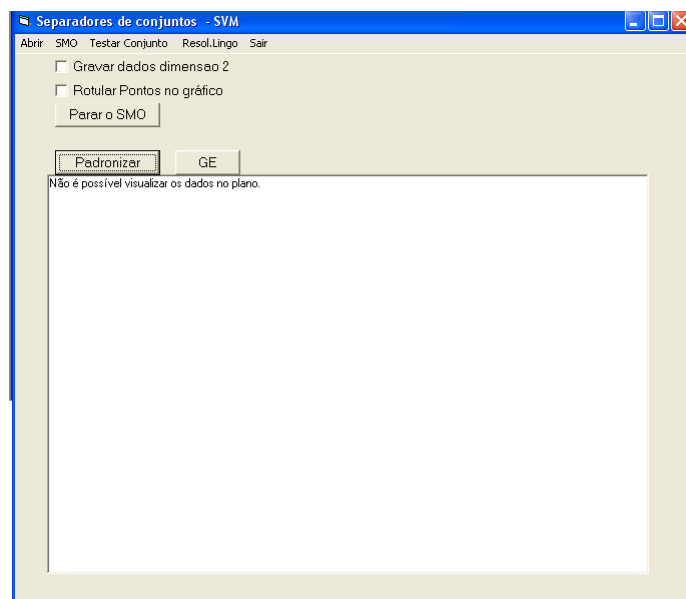


FIGURA 4.2: PÁGINA INICIAL DO PROGRAMA
FONTE: Ales (2008)

O arquivo contendo os dados deve estar no formato texto (.txt). Para isso, foram transformados os dados da planilha eletrônica para o formato mencionado. O valor da variável para as empresas adimplentes continua sendo 1, e para as inadimplentes foi transformado para -1.

Para alimentar o programa, seguem os passos: abrir (abre uma caixa para a escolha do arquivo) e escolher o arquivo com os dados; automaticamente, o programa faz a leitura do arquivo, mostrando na interface o número de pontos (199) e dimensão (21), conforme Figura 4.3.

¹⁹ O *Microsoft Visual Basic* é uma linguagem para desenvolvimento de aplicações visuais para ambiente *Windows* baseado em *Basic* (*Beginners Allpurpose Symbolic Instruction Code*).

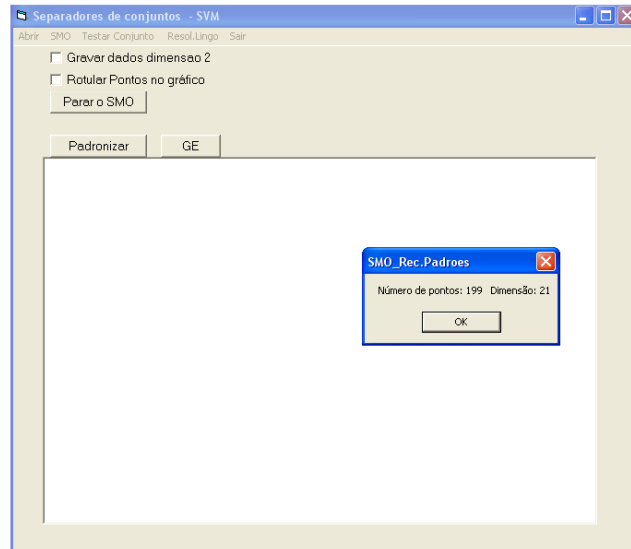


FIGURA 4.3: TELA MOSTRANDO O NÚMERO DE PONTOS E A DIMENSÃO
 FONTE: A autora (2009)

Quando o programa carrega o banco de dados, pode ser iniciada a escolha dos parâmetros da modelagem dual do SVM através do lagrangeano. O usuário fornece a função *kernel*, os parâmetros do *kernel* escolhido, o parâmetro C e pressiona o botão **Confirmar a escolha e continuar** para que o programa inicialize o SMO, como mostra a Figura 4.4 (ALES, 2008):

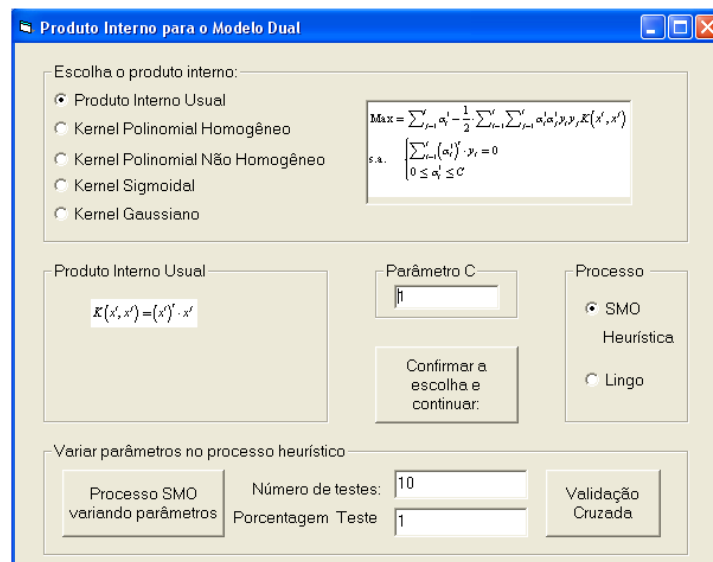


FIGURA 4.4: TELA DE ESCOLHA DA FUNÇÃO *KERNEL* E DOS PARÂMETROS
 FONTE: Ales (2008)

Caso o usuário queira avaliar todos os parâmetros e todas as funções *kernel*, não necessita escolher nenhum parâmetro, apenas pressionar o botão **Processo SMO variando parâmetros**. Automaticamente, o programa executará o algoritmo SMO (ALES, 2008).

Se o objetivo for avaliar os parâmetros escolhidos, o usuário deve escolher o *kernel*, os parâmetros, o número de testes que deseja realizar e a porcentagem de dados que deve ser reservado para testes, inicializando o processo no botão **Validação Cruzada** (ALES, 2008).

4.7.1 Testando diferentes *kernels*

Inicialmente, para o treinamento, foi utilizado o botão **Variando Parâmetros**, onde os valores dos parâmetros e das funções *kernel* mudavam sem a interferência do usuário. Também foi usado todo o conjunto, sem a separação no conjunto de treinamento e teste. Alguns resultados estão descritos na tabela 4.1, onde VS representa o número de vetores suporte não *bounds*, VSB representa o número de vetores suportes *bounds* e porcentagem de erro representa o erro durante o treinamento com os 199 dados.

TABELA 4.1: PERFORMANCE DE ALGUMAS FUNÇÕES *KERNEL* E SEUS PARÂMETROS

Pontos treinados	VS	VSB	Pontos classificados errados	Porcentagem de erro	C	Função <i>kernel</i>	Grau do polinômio	Constante	kapa	Sigma
199	15	12	3	1,51%	1	Produto interno	3	0.75	0	0
199	6	125	64	32,16%	1	Sigmoidal	0	0	0.001	0
199	0	128	0	0,00%	1	Sigmoidal	0	0	10	0
199	0	128	0	0,00%	1	Sigmoidal	0	-0.25	10	0
199	0	128	0	0,00%	1	Sigmoidal	0	-0.5	10	0
199	0	128	0	0,00%	1	Sigmoidal	0	-0.75	10	0
199	0	128	0	0,00%	1	Sigmoidal	0	-1	10	0
199	0	128	0	0,00%	1	Sigmoidal	0	0	100	0
199	0	128	0	0,00%	1	Sigmoidal	0	-0.25	100	0
199	0	128	0	0,00%	1	Sigmoidal	0	-0.5	100	0
.
.
199	24	0	0	0,00%	10	Polinomial Não Homogêneo	2	0	0	0
199	24	0	0	0,00%	10	Polinomial Não Homogêneo	2	0.25	0	0
199	24	0	0	0,00%	10	Polinomial Não Homogêneo	2	0.5	0	0
199	24	0	0	0,00%	10	Polinomial Não Homogêneo	2	0.75	0	0
199	23	0	0	0,00%	10	Polinomial Não Homogêneo	2	1	0	0
.
.
199	22	0	0	0,00%	10	Polinomial Não Homogêneo	3	0.5	0	0
199	199	0	0	0,00%	1000	Gaussiano	0	0	0	0.001
199	198	0	0	0,00%	1000	Gaussiano	0	0	0	0.01
199	186	0	0	0,00%	1000	Gaussiano	0	0	0	0.1

FONTE: A autora (2009)

O programa gera um arquivo de resultados contendo: tipo de SVM, tipo de *kernel*, número de classes, total de vetores suporte, valor da função dual, quantidade de iterações, pesos w e outros. A Figura 4.5 ilustra parte do arquivo de resultado. Essa Figura foi montada apenas a título de ilustração com partes do arquivo para mostrar quais informações ficam disponíveis nos arquivos de resultados.

```

Arquivo: DadosBancarios.txt
17/7/2009 15:13:55
Número de iterações: 7950 Dimensão: 21 Número de Pontos: 157
Número de Vetores Suportes: 18
Número de Vetores Suportes bound: 0
B: 5.650972
W( 1 )=0.2864431
W( 2 )=6.901912E-02
W( 3 )=-0.1476713
W( 4 )=3.200804E-02
W( 5 )=3.053625E-02
W( 6 )=-3.463349E-02
W( 7 )=7.569193E-03
W( 8 )=-0.0425215
W( 9 )=-2.732158E-02
W( 10 )=1.985689E-02
W( 11 )=8.319568E-02
W( 12 )=0
W( 13 )=2.241206E-02
W( 14 )=-0.144727
W( 15 )=3.015265E-02
W( 16 )=-4.618006E-02
W( 17 )=-1.122089E-02
W( 18 )=0.1164234
W( 19 )=1.129723E-02
W( 20 )=2.448419E-02
W( 21 )=1.731382E-02
Função objetivo dual: 3.929523
Número de pontos classificados errados: 0
Porcentagem de erro: 0.000%
Parâmetros C: 10 Kernel: Polinomial Não Homogêneo
Grau do Polinômio: 2 Constante: 5
Kapa: 0 Sigma: 0
Tempo de execução do SMO: 7.14025 segundos
17/7/2009 15:13:55
Número de pontos de teste: 42
Pontos classificados incorretamente:
Número de pontos na fronteira : 0
Número de pontos classificados incorretamente: 0

```

FIGURA 4.5: PARTE DO ARQUIVO RESULTANTE APÓS O SVM
FONTE: A autora (2009)

4.7.2 Resultados obtidos

As diferentes funções *kernel*, variando os parâmetros, obtiveram resultados satisfatórios. Então, foram escolhidas as funções que obtiveram desempenho de erro de treinamento de 0% com todo o conjunto de dados.

Os dados foram novamente treinados, porém, dividindo o conjunto em dois subconjuntos, o de treinamento e o de teste, escolhendo as funções *kernel* e os parâmetros obtidos na tabela 4.2.

Para reforçar os resultados obtidos nesta etapa, o programa disponibiliza a estratégia validação cruzada (*cross validation*). O usuário escolhe o número de testes e a porcentagem do teste. O número de teste escolhido foi $k=10$, conforme recomenda Haykin (1999), e a porcentagem do teste foi a mesma utilizada nos outros modelos, ou seja, 20%.

TABELA 4.2: FUNÇÕES *KERNEL* TREINADAS E SEUS PARÂMETROS

Trein	VS	VSB	C	<i>kernel</i>	Grau	Cte	kapa	sigma	trein	teste
157	20	0	10	Polinomial Não Homogêneo	2	5	0	0	157	42
157	0	100	1	Sigmoidal	0	0	10	0	157	42
157	19	18	1	Gaussiano	0	0	0	1	157	42
157	12	0	10	Polinomial Não Homogêneo	6	0	0	0	157	42
157	14	0	10	Polinomial Não Homogêneo	6	1	0	0	157	42
157	12	5	10	Gaussiano	0	0	0	1	157	42
157	4	14	100	Gaussiano	0	0	0	10	157	42
157	16	1	100	Produto interno	0	0	0	0	157	42
157	12	0	10	Polinomial Homogêneo	5	0	0	0	157	42
157	0	100	20	Sigmoidal	0	-10	0.1	0	157	42
157	19	0	8	Polinomial Homogêneo	3	0	0	0	157	42
157	22	0	3	Polinomial Não Homogêneo	8	2	0	0	157	42
157	1	98	10	Sigmoidal	0	-15	0.8	0	157	42
157	13	41	10	Sigmoidal	0	-2	0.1	0	157	42
157	3	99	1	Gaussiano	0	0	0	15	157	42
157	13	0	100	Polinomial Homogêneo	5	0	0	0	157	42
157	13	0	100	Polinomial Não Homogêneo	6	1	0	0	157	42
157	17	21	1000	Sigmoidal	0	-5	0.2	0	157	42

FONTE: A autora (2009)

TABELA 4.3: PERFORMANCE DAS FUNÇÕES *KERNEL*, DESVIO PADRÃO E TEMPO COMPUTACIONAL

Média erro trein.	Desvio padrão	<i>kernel</i>	Média erro teste	Desvio padrão	Média do Modelo	Tempo computacional (s)
0,00%	0,00%	Polinomial Não Homogêneo	1,49%	1,77%	0,74%	78,44
0,00%	0,00%	Sigmoidal	0,00%	0,00%	0,00%	1,24
1,59%	0,34%	Gaussiano	2,38%	0,00%	1,98%	54,02
1,34%	1,25%	Polinomial Não Homogêneo	4,05%	2,52%	2,69%	5,30
2,10%	1,90%	Polinomial Não Homogêneo	4,52%	5,55	3,31%	4,75
2,04%	0,84%	Gaussiano	4,05%	2,52%	3,04%	51,08
3,82%	1,88%	Gaussiano	3,33%	1,23%	3,57%	3,69
0,13%	0,27%	Produto interno	2,81%	2,79%	1,47%	73,56
1,02%	1,00%	Polinomial Homogêneo	2,86%	3,33%	1,94%	6,36
31,85%	0,00%	Sigmoidal	33,33%	0,00%	32,59%	0,67
0,32%	0,81%	Polinomial Homogêneo	0,95%	1,66%	0,63%	66,77
0,70%	0,92%	Polinomial Não Homogêneo	2,14%	2,37%	1,42%	1,80
31,85%	0,00%	Sigmoidal	33,33%	0,00%	32,59%	0,97
1,47%	0,43%	Sigmoidal	2,38%	0,00%	1,92%	14,80
31,85%	0,00%	Gaussiano	33,33%	0,00%	32,59%	1,78
1,78%	1,30%	Polinomial Homogêneo	3,13%	0,00%	2,45%	5,77
0,96%	1,57%	Polinomial Não Homogêneo	2,38%	3,37%	1,67%	5,28
0,76%	0,40%	Sigmoidal	0,71%	1,15%	0,73%	54,50

FONTE: A autora (2009)

O tempo computacional das funções destacadas foi satisfatório, com 1,24, 66,77 e 54,50 segundos, respectivamente.

Na tabela 4.3, verifica-se que as funções *kernel* destacadas obtiveram os melhores desempenhos dentre as funções treinadas, pois as taxas de erro de treinamento foram de 0%, 0,32% e 0,76%, com desvio padrão de 0%, 0,81% e 0,40%, respectivamente. As taxas de erro de teste foram de 0%, 0,95% e 0,71%, respectivamente com desvio padrão de 0%, 1,66% e 1,15%, respectivamente.

Observa-se também na tabela 4.3 que as outras funções *kernel* (gaussiano e polinomial não homogêneo) também possuem baixas taxas de erros, com desvio padrão pequeno e com poucos vetores suporte.

Verifica-se na tabela 4.2 que o número de vetores de suporte *bound* e número de vetores suporte não *bound* são iguais a 0 e 100 para o *kernel* sigmoidal, causando *overfitting* ou um superajustamento do modelo, não generalizando a superfície de separação, podendo desconsiderá-lo para comparação. O número de vetores de suporte para o *kernel* polinomial homogêneo é 19 (VS) e 0 (VSB), e para o *kernel* sigmoidal 17 (VS) e 21(VSB).

O número de vetores da função *kernel* polinomial homogêneo é aproximadamente 10% do total dos dados de treinamento, e isso gerou uma superfície de separação bem generalizada. Observando os resultados da tabela 4.3, o *kernel* polinomial homogêneo pode alcançar a média de erro de 0,99% na etapa de classificação. Isto representa 99,01 % de acerto do modelo, sendo eleito o *kernel* polinomial homogêneo, com parâmetros $C=8$ e grau do polinômio 3, para realizar a comparação.

4.8 COMPARAÇÃO ENTRE AS TÉCNICAS

Os resultados obtidos nas três técnicas de predição permitiram a comparação entre elas e a análise da melhor técnica.

Os percentuais de acerto nos modelos de regressão logística e redes neurais mostraram-se apropriadas na classificação, atingindo 100% e 93,39% na fase de treinamento, e 96,99% e 90,16% na fase de teste, respectivamente.

A técnica *Support Vector Machine* (SVM's) apresentou resultados bastante próximos entre a fase o treinamento (99,68%) e de teste (99,05%).

QUADRO 4.8: COMPARAÇÃO ENTRE AS TÉCNICAS

	Média de Acertos		
	Treinamento	Teste	Geral
Regressão Logística	100,00%	96,99%	98,50%
Redes Neurais	93,39%	90,16%	91,78%
SVM	99,68%	99,05%	99,37%

FONTE: A autora (2009)

O quadro 4.8 mostra os resultados de classificação obtidos pelos três métodos. Todos eles apresentaram bons resultados de classificação, pois, segundo Picinini *et al.* (2003): “Modelos de *credit scoring* com taxas de acerto acima de 65% são considerados bons por especialistas”.

Desta forma, a técnica eleita como a mais indicada para a predição de empresas adimplentes e inadimplentes é a técnica *Support Vector Machine*, que proporcionou a média de resultados de 99,37%. Contudo, deve ser ressaltado, mais uma vez, que a adoção de qualquer um dos modelos traria bons resultados à instituição financeira.

4.9 TREINAMENTO COM ANÁLISE FATORIAL

Após o primeiro treinamento, foi realizada uma Análise Fatorial (AF) através da Análise de Componentes Principais (ACP), técnica multivariada utilizada para reduzir a dimensionalidade do problema e que encontra-se detalhada no Apêndice B.

4.9.1 Redução de dimensionalidade

Segundo Grobe (2005), a redução de dimensionalidade pode ser efetuada com três objetivos distintos: reduzir ruídos, extrair informações redundantes e facilitar a informação.

O processo de redução é utilizado quando o número de variáveis estudadas é muito grande. Para reduzir a dimensão do vetor original, pode-se criar um novo vetor cujas componentes são combinações lineares das variáveis originais.

A redução de dimensionalidade do vetor de variáveis originais pode ser feita segundo dois critérios (DRAPER e SMITH, 1998):

- Conhecimento informal do especialista utilizando a própria vivência em torno do problema, ou seja, conhecimento empírico;
- Conhecimento científico através dos critérios estatísticos, conhecidos como técnicas de seleção de variáveis.

A criação de um novo vetor de variáveis de menor dimensão, cujos componentes são combinações lineares das variáveis originais, pode ser conduzida via uma técnica de análise multivariada, conhecida como Análise dos Componentes Principais (ANDERSON, 1984).

Outro critério interessante e muito utilizado para reduzir a dimensionalidade dos problemas é a aplicação da Análise Fatorial, que além de poder utilizar o critério dos componentes principais para estimar os pesos fatoriais, possibilita a visualização de grupos que agregam variáveis e são chamados de fatores, oferece o cálculo de comunalidades²⁰ e os valores residuais.

No Apêndice B encontra-se a descrição do Método das Componentes Principais e a Análise Fatorial.

Para a implementação das técnicas de redução de dimensionalidade, utilizou-se a Análise Fatorial através do *software Statistica*, que possibilitou vários cálculos e conclusões, conforme descrito na sequência.

4.9.2 Autovalores e variância explicada

A matriz utilizada para a Análise Fatorial foi a matriz com todos os dados, conforme o quadro 4.9. Porém, quando o programa foi alimentado com a matriz para os cálculos dos autovalores, foi necessário excluir a variável 12, pois não possuía variância.

Das 21 variáveis foi eliminada a variável 12 e a matriz de dados ficou com dimensão 199 x 20, descrita no quadro 4.9.

²⁰ Porção da variância de uma variável X, distribuída entre fatores comuns.

QUADRO 4.9: VARIÁVEIS RESULTANTES

Variável	Informações
X1	Existência de restrições em nome da empresa
X2	Existência de restrições baixadas nos últimos cinco anos em nome da empresa
X3	Tempo de conta no Banco do Brasil
X4	Setor de Atividade
X5	Tempo de Atividade
X6	Número de funcionários
X7	Sede de empresa (imóvel)
X8	Nome do Bairro
X9	Principais clientes
X10	Faturamento bruto anual
X11	Cliente em outro banco
X12	Bens móveis
X13	Seguro Empresarial
X14	Aplicações financeiras no Banco do Brasil
X15	Vendas a prazo
X16	Experiência de crédito no Banco do Brasil
X17	Histórico da conta corrente
X18	Sócios da empresa possuem restrições
X19	Sócios da empresa tiveram restrições baixadas nos últimos cinco anos
X20	Sociedade entre cônjuges

FONTE: A autora (2009)

Um dos cálculos desenvolvidos no *software Statistica* foi para encontrar os autovalores através da Análise de Componentes Principais, conforme indica o quadro 4.10.

QUADRO 4.10: AUTOVALORES E % DE VARIÂNCIA EXPLICADA

Autovalores				
Extração dos principais componentes				
	Autovalor	Variância % total	Autovalor Acumulado	% Variância Explicada
1	4,393195	21,965975	4,393195	21,965975
2	2,4140394	12,070197	6,8072344	34,036172
3	1,8621386	9,3106932	8,669373	43,346865
4	1,7712491	8,8562457	10,440622	52,203111
5	1,3328228	6,6641139	11,773445	58,867225
6	1,0038759	5,0193793	12,777321	63,886604
7	0,9610286	4,8051432	13,738349	68,691747
8	0,8552913	4,2764563	14,593641	72,968204
9	0,8443412	4,2217059	15,437982	77,18991

FONTE: A autora (2009)

Esse quadro indica a presença de 9 autovalores maiores que 0,8²¹. A última coluna do quadro 4.10 reforça a idéia que 9 variáveis são as principais, dentre as 20 iniciais, pois explicam aproximadamente 80% da variância total. As variáveis de menor importância foram identificadas analisando os autovalores. São elas: X3, X1, X19, X6, X14, X16, X15, X17, X18, X13 e X20.

²¹ Kaiser (1960) considera razoável, autovalores superiores a 1,0 como sendo os principais (CUNICO, 2005).

4.9.3 Pesos Fatoriais e Escores Fatoriais

Outro fator de suma importância que o pacote de Análise Fatorial do *Statistica* oferece é o cálculo dos pesos fatoriais que permite distinguir quais fatores entre os 9 estão carregados²² com quais variáveis. Para o cálculo dos pesos fatoriais, optou-se por uma rotação dos fatores, também conhecida como Rotação Varimax descrita no Apêndice B. A rotação permite que os fatores sejam translacionados próximos de variáveis que os carregam com maior intensidade. O quadro 4.11 mostra os pesos fatoriais rotacionados.

QUADRO 4.11: PESOS FATORIAIS ROTACIONADOS

Pesos Fatoriais Rotacionados									
	Fator	Fator	Fator	Fator	Fator	Fator	Fator	Fator	Fator
	1	2	3	4	5	6	7	8	9
X1	0,859575472	0,138383865	0,049410687	0,082161452	0,005327707	0,090366631	-0,143800093	-0,0136788	0,0394792
X2	0,82638161	-0,098472574	-0,061047987	-0,063592818	0,046874106	-0,002683058	0,004964445	0,0929066	-0,1370564
X3	0,130313267	-0,705653823	0,054383717	0,048921841	-0,109833304	0,238463815	-0,131723817	0,533897	0,0343249
X4	-0,019057146	-0,189649668	-0,761309622	0,346872908	-0,072771118	0,02266355	-0,085145045	-0,1098224	-0,035579
X5	0,032565564	0,939970584	-0,005226647	0,036702506	0,090066115	-0,087500038	0,054206292	0,0129569	0,04294
X6	0,18330062	-0,178554294	-0,279375471	-0,056637857	-0,058039374	0,759799625	0,23282627	0,1073997	-0,037806
X7	-0,172378555	0,116673197	0,04596302	0,053907448	0,06294815	0,03474075	0,907272061	-0,0275923	0,1113347
X8	0,010304577	0,101278465	-0,438749257	-0,074994393	0,180051546	-0,075039355	0,256287039	-0,0528426	0,6013022
X9	0,068971687	0,158529066	-0,797323543	-0,175265372	0,043897962	0,154033319	0,008265131	0,0416649	0,0807878
X10	0,055191461	-0,061502899	0,029314669	0,001772264	-0,066241695	0,896549563	-0,120435927	0,0375106	0,0746184
X11	0,080744094	0,079268126	-0,063420343	-0,057362416	0,919033675	-0,071799098	0,027369258	-0,0318144	-0,009574
X12	0,047370909	-0,069428469	0,032255125	0,098449834	-0,032861324	0,067379726	-0,000467247	0,9559005	0,023113
X13	-0,358913106	0,239538848	0,231538934	0,37691504	0,509555186	-0,131733387	0,11899693	-0,0935761	0,1721057
X14	-0,543898087	0,278026324	0,239200301	0,131399868	0,236936908	-0,283276235	0,024078237	-0,0712709	0,1688276
X15	-0,010452236	0,062797156	0,11987866	-0,829404479	0,098654139	0,109278314	-0,003606528	-0,0515989	0,108849
X16	0,003117952	0,077975266	0,071800562	0,837998538	0,096370323	0,072642532	0,032872257	0,0720767	0,1087087
X17	-0,664389542	0,311284732	-0,002118273	-0,010887619	0,268515446	-0,015133186	-0,082991373	0,0881236	0,0309108
X18	0,744498975	0,120908411	0,133701843	0,04922933	0,13145063	0,086670905	-0,365929238	-0,0294192	0,0530103
X19	0,772040671	0,064606277	-0,125112592	-0,041731013	0,073896359	0,049139377	0,033776997	0,1114088	-0,2779263
X20	-0,270412819	-0,011527817	0,05605724	0,043303362	-0,041565559	0,085349933	0,003788406	0,0618058	0,7786822

FONTE: A autora (2009)

O quadro 4.11 acima mostra os pesos fatoriais já rotacionados com destaque para as variáveis mais carregadas. A análise deste quadro mostra claramente o carregamento do fator 1 pelas variáveis 1, 2, 18 e 19. O segundo fator está carregado pelas variáveis 3 e 5, enquanto a variável 14 não carrega nenhum fator com peso acima de 0,7.

Também foram calculadas as comunalidades (Apêndice E) e a matriz dos resíduos (Apêndice F), que, conforme citado no Apêndice B, também são utilizados para certificação de que o modelo fatorial está próximo da realidade.

²² O carregamento em uma variável é indicado pelo peso fatorial. Pesos maiores que 0,7 indicam bom carregamento, ou seja, o fator específico está sendo explicado por essa variável.

Finalizando a aplicação da Análise Fatorial (AF), foram calculados os escores fatoriais após a Rotação Varimax. São esses escores que substituem a matriz modificada (quadro 4.1) do formato 199 x 20, por uma nova matriz, reduzida para 199 X 9, ou seja, 199 indivíduos e 9 escores fatoriais. Portanto, os novos escores fatoriais são os novos dados. A tarefa de redução de dimensionalidade do problema foi concluída.

Parte dessa nova matriz segue no quadro 4.12, e ela completa pode ser encontra no Apêndice G.

QUADRO 4.12: ESCORES FATORIAIS ROTACIONADOS

Escore Fatoriais									
Rotação Varimax Normalizada									
Extração de Componentes Principais									
	Fatores								
	1	2	3	4	5	6	7	8	9
1	0,6648909	-0,9744874	0,452728	1,3818269	-0,4612365	0,9457316	-0,0065582	13,416875	0,3244105
2	0,2211648	-0,1458321	0,9307337	0,6601532	0,1446351	10,444331	-0,3918243	-0,6234942	-0,0651416
3	0,3802346	-1,1158318	1,0883687	-0,6296805	-0,808099	0,0248636	-0,6974102	0,0824364	0,8854135
4	-0,1359353	-0,3543886	1,0106978	-0,5958725	-1,1332435	0,289969	-0,8422428	-0,5361468	0,7158645
5	1,0359352	-0,5654568	0,7327995	-0,7544215	-1,0349225	-0,4302901	-0,3200834	-0,1641656	0,3155412
6	0,3637842	0,2508811	1,0268309	1,9548039	-0,6615866	-0,2318859	-0,4991378	-0,1119284	0,3811072
7	1,0281594	-1,0858567	0,8378237	-0,7254441	-0,8778941	0,3697409	-0,4433574	0,1843634	0,3997112
8	1,13198	-0,7622662	0,4722215	0,3516584	-1,3125384	-0,6442625	-0,2563917	-0,0897996	0,0489969
9	-0,2289961	0,849437	1,1687142	-0,3602353	0,8555836	0,1441265	0,7487109	-0,1773412	0,3251966
10	-0,4448096	0,4538374	1,1223135	-0,2891876	-0,4834922	-0,0796323	-1,091833	-0,2276396	0,6674474
.
.
.
195	-1,4620205	-0,0327684	0,8303788	-0,4920498	-0,6449246	0,1742875	1,6992073	0,0644924	-0,3158159
196	-1,552539	-0,428703	-2,5523839	1,784851	-0,4848291	-0,516791	-0,4259106	-0,2352324	0,6264676
197	-1,6021151	0,1088796	0,6055177	-0,5637767	-0,6399606	-0,0998974	-0,2149395	-0,0173362	-0,260101
198	-1,1591209	0,8917841	0,0675259	0,4927352	-1,0832415	-0,5177716	1,9958825	-0,1177594	0,4467003
199	-0,5608891	0,8493107	0,8889296	0,5979302	-0,3075436	0,0423237	-0,5912568	-0,260537	-0,0660441

FONTE: A autora (2009)

4.9.4 Matriz de treinamento e teste

Após a redução de dimensionalidade, foi preciso separar a matriz em dois conjuntos novamente: o conjunto de treinamento e o conjunto de teste. A proporção é de 20% para o conjunto de teste e 80% para o conjunto de treinamento.

4.10 TÉCNICAS UTILIZADAS

As técnicas utilizadas após a Análise Fatorial foram às mesmas, escolhendo para as redes neurais e *Support Vector Machine* os mesmos parâmetros e funções.

4.10.1 Regressão Logística

Para a implementação da técnica Regressão Logística, utilizou-se novamente o *software Statistica*.

4.10.2 Estimativa dos Parâmetros

Os parâmetros $\beta_0, \beta_1, \beta_2, \dots, \beta_{21}$ foram estimados pelo próprio pacote *Statistica*, que para tanto se utiliza do método da máxima verossimilhança, num contexto de mínimos quadrados não-lineares, aplicando-se o Algoritmo de Quase-Newton.

Os resultados dos parâmetros seguem no quadro 4.13.

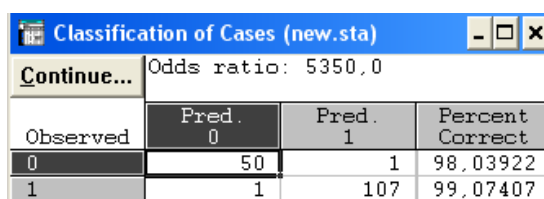
QUADRO 4.13: BETAS ESTIMADOS PELO MÉTODO DA MÁXIMA VEROSSIMILHANÇA - AF

Modelo de Regressão Logística									
Parâmetros estimados									
β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
3,299	10,014	-0,285	0,856	0,258	0,039	2,893	-2,451	0,332	1,108

FONTE: A autora (2009)

4.10.3 Classificações Corretas

A matriz de confusão contendo os acertos e os erros na fase de treinamento é gerada no próprio pacote *Statistica* e segue na figura 4.6, a título de ilustração.



Classification of Cases (new.sta)			
Continue...		Odds ratio: 5350,0	
Observed	Pred. 0	Pred. 1	Percent Correct
0	50	1	98,03922
1	1	107	99,07407

FIGURA 4.6: MATRIZ DE CONFUSÃO GERADA PELO *SOFTWARE STATISTICA* - AF
FONTE: A autora (2009)

4.10.4 Equações da Logit Estimada e do Modelo Estimado

De posse dos parâmetros betas estimados na análise de regressão, foi possível confeccionar as equações do modelo estimado e da logit estimada que podem ser agora utilizadas para realizar a fase de teste. As equações seguem abaixo.

Logit estimada:

$$\hat{g}(x) = 3,299 + 10,014x_1 - 0,285x_2 + 0,856x_3 + \dots + 0,332x_8 + 1,108x_9$$

Modelo Estimado

$$\hat{\pi}(x) = \frac{e^{3,299+10,014x_1-0,285x_2+0,856x_3+\dots+0,332x_8+1,108x_9}}{1 + e^{3,299+10,014x_1-0,285x_2+0,856x_3+\dots+0,332x_8+1,108x_9}}$$

Cada linha da matriz de teste com suas respectivas variáveis é substituída em $\hat{\pi}(x)$, possibilitando o cálculo da variável binária que indicará se o indivíduo em estudo pertence a população “0” ou “1”.

4.10.5 Resultados da Regressão Logística

Depois de estimados os parâmetros, novamente foi testado o modelo. O teste foi realizado utilizando uma planilha eletrônica e o quadro 4.13 indica os resultados parciais da matriz de teste, utilizando os parâmetros estimados na regressão e a função logística. Apresenta também o resultado da função logit estimada e do modelo estimado.

QUADRO 4.13: RESULTADOS PARCIAIS DA TÉCNICA DE REGRESSÃO LOGÍSTICA - AF

Número	Logit	Modelo Estimado	Realidade	Resultado
1	6,9597075	0,999051524	1	Correto
2	15,496278	0,999999814	1	Correto
3	16,213853	0,999999909	1	Correto
4	0,8531818	0,701234108	1	Correto
5	9,7424439	0,999941266	1	Correto
6	15,688829	0,999999846	1	Correto
.
.
.
10	3,6012656	0,973435724	1	Correto
11	8,4929057	0,999795124	1	correto
12	9,6373992	0,999934762	1	correto
13	-1,329462	0,209248387	1	errado
14	12,737342	0,999997061	1	correto
15	11,164585	0,999985833	1	correto
16	0,0490597	0,412262473	1	errado
17	9,4748218	0,999923245	1	correto
18	-16,34601	7,961970E-08	1	errado
.
.
.
38	-12,48381	3,78749E-06	0	correto
39	-14,35387	5,83707E-07	0	correto
40	-0,244037	0,439291735	0	correto

FONTE: A autora (2009)

4.10.6 Resultado final da Regressão Logística - AF

Os resultados finais da técnica de Regressão Logística após a AF estão demonstrados no quadro 4.14.

QUADRO 4.14: RESULTADO FINAL DA REGRESSÃO LOGÍSTICA - AF

Regressão Logística					
Média acerto treinamento	Desvio padrão	Média acerto teste	Desvio padrão	Média de acerto do modelo	Tempo computacional (s)
97,82%	1,38%	94,54%	4,01%	96,18%	menor que 0,1

FONTE: A autora (2009)

Observa-se no quadro 4.14 que o modelo teve desempenho de 97,82% na fase de treinamento e desvio padrão de 1,38%, e na fase de teste teve 94,54% de desempenho e desvio padrão de 4,01%. Logo, a média de acerto do modelo é de 96,18%.

O tempo computacional mostrou-se satisfatório pois, nos 10 conjuntos treinados, o tempo foi menor que 0,1 segundos.

4.11 TÉCNICA REDES NEURAIIS

Para a implementação da técnica de Redes Neurais, novamente foi utilizado o *software MATLAB – Neural Networks Toolbox*.

Para efetuar a verificação dos resultados dos conjuntos de Teste das Redes Neurais, foi elaborada uma planilha eletrônica.

4.11.1 Resultados Obtidos Redes Neurais

O quadro 4.15 demonstra os resultados obtidos com a técnica de Redes Neurais:

QUADRO 4.15: PERFORMANCE DAS REDES TREINADAS - AF

Rede	Média acerto treinamento	Desvio padrão	Média acerto teste	Desvio padrão	Número de neurônios	Número de iterações	Função	Tempo computacional (s)
1	84,90%	12,86%	77,22%	14,60%	6	30	logit	menor que 0,1
2	83,46%	19,51%	80,70%	18,60%	6	50	logit	menor que 0,1
3	87,66%	14,03%	82,45%	12,63%	6	100	logit	menor que 0,1
4	82,91%	18,71%	81,04%	18,18%	8	30	logit	menor que 0,1
5	90,11%	4,20%	82,45%	11,37%	8	50	logit	menor que 0,1
6	88,43%	6,14%	85,57%	9,25%	8	100	logit	menor que 0,1
7	79,48%	18,95%	77,22%	18,65%	10	100	logit	menor que 0,1
8	89,11%	8,00%	83,77%	12,51%	12	100	logit	menor que 0,1

FONTE: A autora (2009)

No quadro 4.15, pode-se destacar a rede 6, que possui acerto de 88,43% e 85,57% e desvio padrão de 6,14% e 9,25%, nas fases de treinamento e teste, respectivamente.

Sendo assim, a rede eleita é a número 6 com média de acerto do modelo de 87,00% e baixo tempo computacional.

4.11.2 Pesos e bias

Seguem no quadro 4.16, a título de ilustração, os pesos e bias alcançados com a rede 6.

QUADRO 4.16: PESOS E BIAS DA REDE Nº 6 - AF

Pesos da camada intermediária: 8 neurônios							
w1 = matriz 8 x 21							
-2,1543	-2,6062	-1,0074	-0,77504	-0,52735	0,27942	0,13228	
-1,4667	2,3547	-1,4153	0,21316	0,11572	0,84832	1,5587	
-1,895	-1,0387	0,46301	0,0032777	-0,088232	1,9579	0,79263	
-5,1081	-0,22417	0,77187	0,25498	-1,0089	-2,4298	0,70915	
-5,1851	-1,4232	-0,013436	-0,10298	0,45551	-0,67894	0,042658	
0,54464	0,20526	1,3238	0,61902	0,078435	0,11038	0,35872	
-3,9426	-0,80644	-1,523	0,21565	0,3947	-1,2787	1,5888	
2,7524	-1,2384	-1,7382	-0,6412	0,42757	-0,11157	-0,19382	
0,8817	0,41049						
0,05705	-0,042824						
0,34132	-0,96018						
3,1042	-0,053728						
0,27348	1,1032						
-0,84727	-0,57976						
0,20286	-3,1565						
-2,0476	-1,3553						
Pesos da camada de saída: 1 neurônio							
w2 = vetor 1 x 8							
3,877	0,0065615	2,1254	-2,8143	-2,6389	0,14	-2,3746	-2,2473
Bias da camada intermediária e da saída							
b1 = vetor 8 x 1		b2	1,8311				
5,0166							
1,8767							
2,9485							
0,84238							
-0,74931							
-3,456							
3,6885							
-5,1564							

FONTE: A autora (2009)

4.12 TÉCNICA *SUPPORT VECTOR MACHINES (SVM's)*

Para a implementação do *Support Vector Machine*, novamente foi utilizado o programa desenvolvido por Ales (2008) em linguagem *Visual Basic*, possibilitando a verificação de todo o processo e a visualização das superfícies de separação quando os dados estiverem no plano cartesiano.

Foram realizados 18 testes, utilizando as mesmas funções *kernel* e parâmetros antes da Análise Fatorial, e os resultados estão demonstrados na tabela 4.3.

TABELA 4.4: FUNÇÕES *KERNEL* TREINADAS E SEUS PARÂMETROS - AF

Trein	VS	VSB	C	<i>kernel</i>	Grau	Cte	kapa	sigma	trein	teste
157	25	8	10	Polinomial Não Homogêneo	2	5	0	0	157	42
157	2	46	1	Sigmoidal	0	0	10	0	157	42
157	44	43	1	Gaussiano	0	0	0	1	157	42
157	26	18	10	Polinomial Não Homogêneo	6	0	0	0	157	42
157	21	6	10	Polinomial Não Homogêneo	6	1	0	0	157	42
157	56	11	10	Gaussiano	0	0	0	1	157	42
157	5	41	100	Gaussiano	0	0	0	10	157	42
157	41	16	100	Produto interno	0	0	0	0	157	42
157	43	8	10	Polinomial Homogêneo	5	0	0	0	157	42
157	3	99	20	Sigmoidal	0	-10	0.1	0	157	42
157	39	9	8	Polinomial Homogêneo	3	0	0	0	157	42
157	19	7	3	Polinomial Não Homogêneo	8	2	0	0	157	42
157	1	91	10	Sigmoidal	0	-15	0.8	0	157	42
157	27	64	10	Sigmoidal	0	-2	0.1	0	157	42
157	4	89	1	Gaussiano	0	0	0	15	157	42
157	37	8	100	Polinomial Homogêneo	5	0	0	0	157	42
157	23	6	100	Polinomial Não Homogêneo	6	1	0	0	157	42
157	28	73	1000	Sigmoidal	0	-5	0.2	0	157	42

FONTE: A autora (2009)

TABELA 4.5: PERFORMANCE DAS FUNÇÕES *KERNEL*, DESVIO PADRÃO E TEMPO COMPUTACIONAL - AF

Média erro trein.	Desvio padrão	<i>kernel</i>	Média erro teste	Desvio padrão	Média do Modelo	Tempo computacional (s)
9,24	2,96%	Polinomial Não Homogêneo	12,86%	7,03%	11,05%	267,43
22,84	3,59%	Sigmoidal	11,39%	3,64%	17,12%	156,21
5,48	1,35%	Gaussiano	5,00%	2,62%	5,24%	74,00
22,87	15,78%	Polinomial Não Homogêneo	22,38%	23,52%	22,62%	69,83
16,18	10,42%	Polinomial Não Homogêneo	17,14%	16,56%	16,66%	20,20
5,35	0,86%	Gaussiano	4,52%	3,06%	4,94%	409,05
25,48	3,34%	Gaussiano	22,62%	4,09%	24,05%	7,53
9,17	1,86%	Produto interno	6,67%	6,02%	7,92%	463,46
3,19	0,90%	Polinomial Homogêneo	2,86%	3,33%	3,02%	101,64
31,85	0,00%	Sigmoidal	33,33%	0,00%	32,59%	7,93
3,89	1,25%	Polinomial Homogêneo	1,90%	2,93%	2,90%	247,14
14,33	9,19%	Polinomial Não Homogêneo	9,05%	10,34%	11,69%	15,03
28,66	10,07%	Sigmoidal	36,67%	10,54%	32,66%	2,88
8,53	0,91%	Sigmoidal	4,76%	0,00%	6,65%	58,98
13,12	4,45%	Gaussiano	13,33%	6,75%	13,23%	62,87
3,44	0,81%	Polinomial Homogêneo	2,14%	2,08%	2,79%	123,19
17,58	10,35%	Polinomial Não Homogêneo	16,90%	12,72%	17,24%	17,70
7,83	1,20%	Sigmoidal	5,00%	0,75%	6,42%	68,59

FONTE: A autora (2009)

O tempo computacional das funções destacadas foi de 101,64, 247,14 e 123,19 segundos, respectivamente.

Na tabela 4.5, verifica-se que as três funções *kernel* polinomiais homogêneo destacadas, possuem taxa média de erro de treinamento de 3,19%, 3,89% e 3,44%, respectivamente, e desvio padrão de 0,90%, 1,25% e 0,81%, respectivamente.

As taxas de erro na fase de teste são 2,86%, 1,90% e 2,14%, com desvio padrão de 3,33%, 2,93% e 2,08%, respectivamente.

Observa-se também na tabela 4.5 que as outras funções *kernel* (gaussiano, produto interno e sigmoidal) também possuem baixas taxas de erros com baixo desvio padrão.

Verifica-se na tabela 4.4 que o número de vetores de suporte *bound* e número de vetores suporte não *bound* para os *kernel*s homogêneo, é 19 (VS) e 0 (VSB), 39 (VS) e 9 (VSB) e 37 (VS) e 8 (VSB), respectivamente.

O número de vetores da função *kernel* polinomial homogêneo ($C=100$) é aproximadamente 22,62% do total dos dados de treinamento, e isso gerou uma superfície de separação bem generalizada. Observando os resultados da tabela 4.5, o *kernel* polinomial homogêneo de grau 5 e $C=100$ alcançou a média de erro de 2,79% na etapa de classificação. Isto representa 97,21 % de acerto do modelo, sendo eleito para realizar a comparação.

4.13 COMPARAÇÃO ENTRE AS TÉCNICAS

Os resultados obtidos nas três técnicas de predição permitiram a comparação entre elas e a análise da melhor técnica.

Na fase de treinamento, as duas técnicas, Regressão Logística e Redes Neurais, mostraram-se apropriadas, atingindo o acerto de 100% e 99,54%, respectivamente, no treinamento. Porém, na fase de teste, a porcentagem de acerto foi de 88,88% para Regressão Logística e de 94,30% para as Redes Neurais.

A técnica *Support Vector Machine* (SVM's) apresentou resultados bastante próximos tanto para o treinamento (96,11%) quanto para o teste (98,10 %). Mesmo após a Análise Fatorial, foi eleita a melhor técnica para o problema proposto.

A comparação entre elas é demonstrada no quadro 4.17.

QUADRO 4.17: COMPARAÇÃO ENTRE AS TÉCNICAS - AF

	Média de Acertos – Análise Fatorial		
	Treinamento	Teste	Modelo
Regressão Logística	97,82%	94,54%	96,18%
Redes Neurais	88,43%	85,57%	87,00%
SVM	96,56%	97,86%	97,21%

FONTE: A autora (2009)

4.14 RESULTADO FINAL/COMPARAÇÃO ENTRE AS TÉCNICAS

QUADRO 4.18: COMPARAÇÃO ENTRE AS TÉCNICAS ANTES E DEPOIS DA AF

	Média de Acertos			Média de Acertos – Análise Fatorial		
	Treinamento	Teste	Modelo	Treinamento	Teste	Modelo
Regressão Logística	100,00%	96,99%	98,50%	97,82%	94,54%	96,18%
Redes Neurais	93,99%	90,16%	91,78%	88,43%	85,57%	87,00%
SVM	99,68%	99,05%	99,37%	96,56%	97,68%	97,21%

FONTE: A autora (2009)

Desta forma, a técnica mais indicada para análise de crédito bancário e classificação de empresas adimplentes e inadimplentes dos dados apresentados neste trabalho, é *Support Vector Machine*, que apresentou melhor porcentagem de acerto nos modelos finais (99,37% e 97,21%).

CAPÍTULO V

5 CONCLUSÕES E SUGESTÕES DE TRABALHOS FUTUROS

Neste capítulo, são apresentadas as principais conclusões obtidas do estudo e algumas sugestões para trabalhos futuros.

5.1 CONCLUSÕES

O objetivo deste estudo foi analisar os modelos de previsão de risco de crédito por meio da comparação das técnicas Regressão Logística, Redes Neurais e *Support Vector Machine*. As características especiais de cada uma das técnicas foram analisadas para encontrar o modelo com melhor desempenho na predição de clientes adimplentes e inadimplentes.

Todos os passos para a obtenção de um modelo de previsão de risco de crédito foram apresentados, desde a definição da população para a aplicação até os passos para implementação dos modelos. O uso de algoritmos matemáticos possibilitou o conhecimento mais profundo dos *softwares MATLAB* e *Statistica*.

As três técnicas apresentadas e testadas neste trabalho: Regressão Logística, Redes Neurais e *Support Vector Machine*, mostraram ser ferramentas de grande valia para os analistas de crédito bancário, obtendo respectivamente 98,50%, 91,78% e 99,37% de acerto na classificação antes da Análise Fatorial.

Para o problema em questão, observou-se que, antes da Análise Fatorial, a porcentagem de acerto na classificação de empresas adimplentes e inadimplentes foi maior do que após a Análise Fatorial (96,18%, 87,00% e 97,21%), apesar de apresentar uma taxa de classificação também satisfatória. Essa técnica, por meio da Análise dos Componentes Principais, foi essencial para enriquecer o trabalho e fornecer conhecimentos na área de análise multivariada, área da Estatística que, com a popularização do computador e o surgimento de novos *softwares*, vêm sendo largamente utilizada em trabalhos acadêmicos.

Desta forma, observou-se que a técnica *Support Vector Machine* teve um melhor desempenho em relação às outras técnicas, tanto antes quanto após a AF, seguido da Regressão Logística e finalmente das Redes Neurais.

Support Vector Machine é uma área de pesquisa que está em franco desenvolvimento, restrita ainda a poucos grupos, porém já alcançando um público maior. Como a teoria é totalmente formalizada, embora ainda haja muito a ser explorado, o enfoque da maioria dos trabalhos são as aplicações. Busca-se um dimensionamento do poder de abrangência de SVM em problemas práticos. Um dos problemas investigados é o reconhecimento de padrões de empresas adimplentes e inadimplentes que pode ser útil ao analista de crédito.

Neste contexto, foram descritos os conceitos básicos a respeito das SVM's para problemas de classificação, os quais também podem ser consultados em Lorena (2007). Com princípios embasados na teoria de aprendizado estatístico, essa técnica se caracteriza por apresentar uma boa capacidade de generalização.

As SVM's também são robustas diante de dados de grande dimensão, sobre os quais outras técnicas de aprendizado comumente obtêm classificadores super ou sub ajustados. Outra característica atrativa é a convexidade do problema de otimização formulado em seu treinamento, que implica na existência de um mínimo global. Essa é uma vantagem das SVM's sobre, por exemplo, as Redes Neurais Artificiais, em que há mínimos locais na função objetivo minimizada. Além disso, o uso de funções *kernel* na não-linearização das SVM's torna o algoritmo eficiente, pois permite a construção de simples hiperplanos em um espaço de dimensão de forma tratável do ponto de vista computacional.

Em comparação com a técnica estatística Regressão Logística cuja preparação das variáveis demora muito tempo e, no caso de muitas variáveis, deve-se fazer uma pré-seleção das mais importantes baseando-se em análises separadas e alguns modelos resultantes são de difícil implementação.

Porém, entre as principais limitações das SVM's, encontra-se a sua sensibilidade a escolhas de parâmetros e a dificuldade de interpretação de modelo gerado por essa técnica.

O presente trabalho também se limitou a apresentar a formulação original do SVM, o qual é capaz de lidar apenas com problemas de classificação binários. Existe uma série de técnicas que podem ser empregadas na generalização das SVM's para a solução de problemas multiclases.

As SVM's também podem ser aplicadas na solução de problemas de regressão e no agrupamento de dados (aprendizado não supervisionado). Contudo, o problema de otimização para o seu treinamento deve ser formulado para lidar com as características e objetivos de cada problema.

Os problemas de aprendizagem de máquina estão ainda em uma fase inicial de desenvolvimento. As SVM's são muito promissoras, baseadas em fortes e não-triviais teorias matemática e estatística.

A técnica SVM mostrou ser uma ferramenta de grande utilidade para auxiliar os analistas de crédito a decidir sobre a sua concessão, pois classificou as empresas em solventes e insolventes com baixo índice de erro.

Esses testes contribuíram para uma compreensão mais prática da técnica, reforçando o conhecimento teórico e confirmando a aplicabilidade do SVM à análise de crédito bancário.

Os bancos e instituições de concessão de crédito, utilizando suas informações cadastrais, têm condições de avaliar novas empresas que estão em busca de crédito.

Não se pode afirmar que a empresa que obteve um sinal favorável ao crédito vai ser adimplente, pois fatores externos como a economia do país, ou problemas inesperados como enchente, incêndio, roubo, podem alterar o comportamento da empresa em razão dos compromissos adquiridos.

A experiência profissional do analista, aliada à técnica SVM, são instrumentos que podem ajudar na tarefa de tomada de decisão. Porém, nunca por si só essas técnicas poderão substituir a figura do especialista no contexto da análise de crédito.

5.2 TRABALHOS FUTUROS

A seguir, estão descritas algumas sugestões que poderão ser objeto de estudos, de modo a aperfeiçoar a metodologia aqui abordada:

- Implementar essas técnicas de reconhecimento de padrões para maior número de casos, na análise de crédito para pessoas físicas e considerando o valor a ser concedido, assim como o prazo de restituição;
- Aprofundamento teórico das funções *kernel*;
- Na tentativa de reduzir a margem de erro, o questionário inicial poderia ser reformulado, aumentando o número de perguntas e incluindo variáveis qualitativas;
- Avaliar e implementar outros métodos de Aprendizado de Máquina na predição de empresas que foram citados na revisão de literatura.
- Fazer outros tipos de testes de validação para os métodos utilizados.

REFERÊNCIAS

- ALES, V. **O Algoritmo Sequential Minimal Optimization de Support Vector Machine: Uma Técnica de Reconhecimento de Padrões**. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná, PR, 2008.
- ALTMAN, E. L. Financial ratios, discriminant analysis and the prediction of corporation bankruptcy. *Journal of Finance*, v.23, n.4, 1968.
- ANDERSON, T. W. *A introduction to multivariate statistical analysis*. Nova York: Miley, 1984.
- BANDYOPADHYAY, S. Source composite curve for waste reduction. *Chemical Engineering Journal*. v.125, p.99-110, 2006.
- BARANAUSKAS, J. A.; MONARD, M. C. Conceitos de aprendizagem de máquinas. In S.O. Rezende, editor, *Sistemas Inteligentes – Fundamentos e Aplicações*, p.89-114. Editora Manole, 2003.
- BERNHARDSEN, E. A model of Bankruptcy Prediction. Thesis (Cand. Oecon.) – University of Oslo, Oslo (Noruega), 2001. Disponível em: www.norges-bank.no/publikasjoner/arbeidsnotater/pdf/arb-2001-10.pdf. Acesso em 09 de jun. 2009.
- BERNI, N. T. **Operação e concessão de crédito: os parâmetros para a decisão de crédito**. São Paulo: Atlas, 1999.
- BERTESEKAS, D. *Nonlinear programming*. Nashua, NH: Athena Scientific, 1995.
- BLATT, A. **Avaliação de Risco e Concessão de Crédito: um enfoque prático**. São Paulo: Nobel, 1999.
- BLATT, A. **Crédito: dicas práticas para analisar e conceder**. São Paulo: STS, 2000.
- BLATT, A. **Dicas para conceder crédito com menos risco**. São Paulo: Nobel, 1998.
- BUGERA, V.; KONNO, H.; URVASEV, S. *Credit cards scoring with quadratic utility function*. Risk Management and Financial Engineering Lab Center for Applied Optimization, Department of Industrial and Systems Engineering, University of Florida, Gainesville, 2002.
- BURDEN, R. L.; FAIRES, J. D. **Análise Numérica**. São Paulo: Pioneira Thomson Learning, 2003.
- BURGESS, C. J. C. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 2(2):1–43, 1998.
- CAOQUETTE, J. B.; ALTMAN, E. I.; NARAYANAN, P. **Gestão do risco de crédito: o próximo grande desafio financeiro**. Rio de Janeiro: Qualitymark, 2000.
- CARVALHO, A. P. L. F. [S1,s.d.] **Redes Neurais Artificiais**. Disponível em <http://www.icmc.usp.br/~andre/research/neural>. Acesso em 29 maio 2009.

CARVALHO, B. P. R. **Novas Estratégias para Detecção Automática de Vetores de Suporte em *Least Squares Support Vector Machine***. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Minas Gerais, MG, 2005.

CHAPELLE, O.; VAPNIK, V.; BOUSQUET, O.; MUKHERJEE, S. Choosing multiple parameters for support vector machines. *Machine Learning*, 46(1-3):131–159, 2002.

CHEESEMAN, P.; STUSTZ, J. Bayesian classification (Autoclass: Theory and results in knowledge Discovery data mining), 1990. Disponível em: <http://ic.arc.nasa.gov/ic/projects/bayes-group/Autoclass-c-program.html>. Acesso em 22 jul. 2009.

CORRAR, L. J. *et al. Análise multivariada: para cursos de Administração, Ciências Contábeis e Economia*. São Paulo: Atlas, 2007.

CUNICO, L. H. B. **Técnicas em *data mining* aplicadas na predição de satisfação de funcionários de uma rede de lojas do comércio varejista**. Dissertação (Mestrado em Métodos Numéricos) – Universidade Federal do Paraná, PR, 2005.

DIAS, U. M. **Predição da função das proteínas sem alinhamento usando máquinas de vetor de suporte**. Dissertação (Mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas, AL, 2007.

DRAPPER, N. R.; SMITH, H. *Applied Regression Analysis*. 2.ed. Nova York: Wiley, 1998.

DUAN, K. S. S.; KEERTHI, S. S.; POO, A. N. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing*, 51:41–59, 2003.

DUDA, R. O.; HART, P. E. *Pattern Classification*. Wiley-Interscience, 2nd edition, 1973.

EIFERT, D. S. **Análise quantitativa na concessão de crédito versus inadimplência: um estudo empírico**. Dissertação (Programa de Pós-Graduação em Administração) – Universidade Federal do Rio Grande do Sul, RS, 2003.

ELIZABETSKY, R. **Um modelo matemático para a decisão no banco comercial** (Trabalho apresentado ao Departamento de Engenharia de Produção da Escola Politécnica da USP), 1976.

FAUSETT, L. **Fundamentals de Neural Networks: architectures, algorithms and applications**. Ed Prentice-Hall. New Jersey, 1994.

FENSTERSTOCK, F. Credit Scoring and the Next Step. *Business Credit*, v. 107, n. 3, p. 46-49, 2005.

FRIEDLANDER, A. **Elementos de Programação Não-Linear**. São Paulo: Editora Unicamp, Série Manuais, 1994.

GONÇALVES, E. B. **Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéticos.** Dissertação (Departamento de Administração) – Universidade de São Paulo, SP, 2005.

GROBE, J. R. **Aplicações da Estatística Multivariada na Análise de Resultados em Experimentos com solo e animais.** Dissertação (Mestrado em Métodos Numéricos) – Universidade Federal do Paraná, PR, 2005.

GUIMARÃES, I. A.; CHAVES NETO, A. Reconhecimento de padrões: metodologias estatísticas em crédito ao consumidor. *RAE – eletrônica* – vol.1, n.1, jul-dez, 2002.

Disponível em:

<http://www.rae.com.br/eletronica/index.cfm?FuseAction=Artigo&ID=1215&Secao=FINANÇAS2&Volume=1&Numero=2&Ano=2002>. Acesso em 02 jun. 2009.

GUIMARÃES, J. B. **Financiamento de Micro e Pequenas empresas em uma instituição pública de créditos.** Minas Gerais: PUC-MG, 2002.

GUJARATI, D. N. **Econometria básica.** 3. ed. São Paulo: Pearson Education, 2000.

HAIR JR., J. F. *et al.* **Multivariate data analysis.** 5. ed. New Jersey: Prentice Hall, 1998.

HAIR JR., J. F. *et al.* **Multivariate Data Analysis.** 6th ed. New Jersey: Pearson Prentice Hall, 2005.

HAYKIN, S. **Neural Networks: a comprehensive foundation.** 2.ed. New Jersey: Prentice-Hall, 1999.

HSU, C. W; CHANG, C. C.; LIN, C. J. A practical guide to Support Vector Classification. Taiwan, 2008. Disponível em: <http://www.csie.ntu.edu.tw/~cjlin>. Acesso em 22 jul. 2009.

HOSMER, D.W.; LEMESHOW, S. **Applied Logistic Regression.** Nova York: Wiley, 1989.

HUANG, Z.; CHEN, H.; HSU, C.; CHEN, W.; WU, S. Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, n.37, p.543-558, 2003.

KANITZ, S. C. Como prever falências de empresas. *Revista Exame*, p.95-102, dez. 1974.

KROSE, B. J. A.; VAN DER SMAGT, P. P. **An Introduction to Neural Networks.** Amsterdam: University of Amsterdam, 1993.

LAI, K. K; YU, L.; WANG, S.; ZHOU, L. Credit risk analysis using a reliability-based neural network ensemble model. *ICANN, Part II*, LNCS 4132, p.682-690, 2006.

LEMO, E. P. **Análise de crédito bancário com o uso de data mining: redes neurais e árvores de decisão.** Dissertação (Programa de Pós-Graduação em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná, PR, 2003.

LEMO, E. P.; STEINER, M. T. A.; NIEVOLA, J. C. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de *data mining*. *Revista de Administração*, São Paulo-SP, v.40,n.3,p.225-234, jul./ago./set. 2005.

LORENA, A.C.; CARVALHO, A. C. P. L. F. Uma introdução às *Support Vector Machine*. *RITA*, V. XIV, Número 2, 2007.

MARTINELI, E. **Extração de conhecimento de Redes Neurais Artificiais**. Dissertação (Mestrado em Ciências) – USP, SP, 1999.

MENDES FILHO, E. F.; CARVALHO, A. C. P. L. F.; MATIAS, A. B. Utilização de redes neurais artificiais na análise de risco de crédito a pessoas físicas. In: III SIMPÓSIO BRASILEIRO DE REDES NEURAIAS, 1996, Recife. **Anais**.

MÜLLER, K. R.; MIKA, S.; RÄTSCH, G.; TSUDA, K.; SCHÖLKOPF, B. An introduction to kernel-based learning algorithms. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 12(2):181–201, Março 2001.

NGYEN, H. N.; OHN, S.Y.; CHAE, S. H.; SONG, D. H.; LEE, I. Optimizing Weighted Kernel Function for Support Vector Machine by Genetic Algorithm. *MICAI*, 2006 Springer-Verlag Berlin Heidelberg, 2006. Disponível em: <http://www.springerlink.com>. Acesso em 10 jul. 2009.

OHTOSHI, C. **Uma comparação de Regressão Logística, Árvores de Decisão e Redes Neurais: Analisando Dados de Crédito**. Dissertação de Mestrado (Departamento de Estatística) – Universidade de São Paulo, SP, 2003.

OSUNA, E.; FREUND, R.; GIROSI, F. Support vector machines: Training and applications. *Technical Report AIM-1602*, 1997a.

PALAZZO, F. Algoritmos para computação evolutiva. Universidade Católica de Pelotas – Escola de Informática. Grupo de Pesquisa em Inteligência Artificial. Disponível em: <http://ia.ucpel.tche.br/~lpalazzo>. Acesso em 23 maio 2009.

PAIVA, C. A. **Administração do Risco de Crédito**. Rio de Janeiro: Qualitymark Ed., 1997.

PICININI, R.; OLIVEIRA, G. M. B.; MONTEIRO, L. H. A. (2003) Mineração de Critério de *Credit Scoring* Utilizando Algoritmos Genéticos. Bauru: SIMPÓSIO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE, 2003, Bauru, SP. **Anais do VI Simpósio Brasileiro de Automação Inteligente**, p.463-466.

PIRES, I. R. N. **Um modelo estratégico para a análise de crédito utilizando redes neurais artificiais**. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Uberlândia, MG, 2008.

PLATT, J. C. Sequential Minimal Optimization: A Fast Algorithm for training Support Vector Machines. Microsoft Research. Technical Report MSR-TR-98-14. April 21, 1998. Disponível em: <http://research.microsoft.com/users/jplatt/smoTR.pdf>. Acesso em 12 jul. 2009.

PROTIL, R. M.; SILVA, W. V. Utilização da função de análise discriminante linear e o modelo de regressão logística na previsão de insolvência de cooperativas agrícolas do Estado do Paraná. ACEITO PARA PUBLICAÇÃO NO XLIV CONGRESSO DA SOBER, Fortaleza-CE, 23-27 Julho de 2006.

RABUSKE, R. A. **Inteligência Artificial**. Florianópolis: Ed. UFSC, 1995.

RAMÍREZ, F. O. P.; CASTAÑO, H. F. Las redes neuronales y la evaluación Del riesgo de crédito. *R. Ingenierías Universidad de Medellin*, v.6, n.10, p.77-91, 2007.

ROSA, P. T. M. **Modelos de Credit Scoring: Regressão Logística, CHAID e REAL**. Dissertação de Mestrado (Departamento de Estatística) – Universidade de São Paulo, SP, 2000.

SMOLA, A. J.; SCHÖLKOPF, B. **Learning with Kernels**. The MIT Press, Cambridge, MA, 2002.

SANTOS, A. M.; SEIXAS, J. M.; PEREIRA, B. B.; MEDRONHO, R. A. Usando Redes neurais artificiais e regressão logística na predição da Hepatite A. *Revista Brasileira de Epidemiologia*, São Paulo, 2005. **Artigo**.

SANTOS, E. M. **Teoria e aplicação de Support Vector Machine à aprendizagem e reconhecimento de objetos baseados na aparência**. Dissertação (Programa de Pós-Graduação em Informática) – Universidade Federal da Paraíba, PA, 2002.

SANTOS, J. O. **Análise de crédito: empresas e pessoas físicas**. São Paulo: Atlas, 2000.

SCARPEL, R. A. Utilização de *support vector machine* em previsão de insolvência de empresas. SBPO – XXXV VII SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, Gramado, RS, 27-30/set., 2005.

SCHRICKEL, W. **Análise de crédito: concessão e gerência de empréstimos**. São Paulo: Atlas, 2000.

SECURATO, J. R. **Crédito – Análise e Avaliação do Risco – Pessoas Físicas e Jurídicas**. São Paulo: Saint Paul, 2002.

SELAU, L. P. R. **Construção de modelos de previsão de risco de crédito**. Dissertação (Programa de Pós-Graduação em Engenharia de Produção) – Universidade Federal do Rio Grande do Sul, RS, 2008.

SHAW-TAYLOR, J.; CRISTIANINI, N. **Kernel Methods and Pattern Analysis**. United Kingdom: Cambridge: 2000.

SHUNIN, Y. N.; KNIGA, N. Neural networks modeling of business situations and decision-making analysis. *Computer Modelling and New Technologies*, v.9, n.2, p.17-26, 2005.

SERVIGNY, A.; RENAULT, O. **Measuring and managing credit risk**. New York: McGraw-Hill, 2004.

SILVA, J. P. Os dois lados do crédito. *GV Executivo*, v.5, n.3, p.68-72, jul.-ago.2006.

SILVA, J. P. **Gestão e Análise de Risco de Crédito**. São Paulo: Atlas, 2003.

STEINER, M. T. A.; CARNIERI, C.; KOPITTKE, B. H.; STEINER NETO, P. J. Sistemas especialistas probabilísticos e redes neurais na análise do crédito bancário. *Revista de Administração*, São Paulo-SP, n.3, p.56-67, julho/setembro 1999.

TATIBANA, C. Y.; KAETSU, D. Y. Uma introdução às redes neurais, 2000. Disponível em: <http://www.din.uem.br/ia/neurais/#links>. Acesso em 10 jul. 2009.

VAPNIK, V. N. *Statistical Learning Theory*. John Wiley and Sons, 1998.

VELLASCO, M.; CORREA, M. F. **Análise de Risco de Crédito em Correspondentes Bancários através de Redes Neurais, 2006.**

WESTGAARD, S.; WIJST, N. V. D. Default probabilities in a corporate bank portfolio: a logistic model approach. *European Journal of Operational Research*, 135, p. 338-349, 2001.

YIN, R. Case study Research: designs and methods. California: Sage Publications Inc., 1984. (*Applied social research methods series*; v.5).

YIN, R. **Estudo de caso: planejamento e métodos**. 3.ed. Porto Alegre: Bookman, 2005.

APÊNDICE A - QUESTIONÁRIO

Código	Informações	Valor
A	Existência de restrições em nome da empresa	1=(<input type="checkbox"/>) Sim 2=(<input type="checkbox"/>) Não
B	Existência de restrições baixadas nos últimos cinco anos em nome da empresa	1=(<input type="checkbox"/>) Sim 2=(<input type="checkbox"/>) Não
C	Tempo de conta no Banco	Valor Numérico em Meses
D	Setor de Atividade	1=(<input type="checkbox"/>) Comércio 2=(<input type="checkbox"/>) Indústria 3=(<input type="checkbox"/>) Serviços
E	Tempo de Atividade	1=(<input type="checkbox"/>) Mais de 9 anos 2=(<input type="checkbox"/>) De 6 a 9 anos 3=(<input type="checkbox"/>) De 3 a 5 anos 4=(<input type="checkbox"/>) De 1 a 2 anos 5=(<input type="checkbox"/>) Menos de 1 ano
F	Número de funcionários	Valor Numérico
G	Sede de empresa (imóvel)	1=(<input type="checkbox"/>) Próprio 2=(<input type="checkbox"/>) Alugado 3=(<input type="checkbox"/>) Cedido
H	Nome do Bairro	1=(<input type="checkbox"/>) Centro 2=(<input type="checkbox"/>) Outros
I	Principais clientes	1=(<input type="checkbox"/>) Pessoas físicas 2=(<input type="checkbox"/>) Pessoas jurídicas 3=(<input type="checkbox"/>) Misto
J	Faturamento bruto anual	Valor Numérico
K	Cliente em outro banco	1=(<input type="checkbox"/>) Sim 2=(<input type="checkbox"/>) Não
L	Bens imóveis	Valor Numérico
M	Bens móveis	Valor Numérico
N	Seguro Empresarial	1=(<input type="checkbox"/>) Sim 2=(<input type="checkbox"/>) Não
O	Aplicações financeiras no Banco	1=(<input type="checkbox"/>) Sim >8.000 2=(<input type="checkbox"/>) Sim 4.000 a 8.000 3=(<input type="checkbox"/>) Sim 2.000 a 4.000 4=(<input type="checkbox"/>) Sim < 2.000 5=(<input type="checkbox"/>) Não
P	Vendas a prazo	1=(<input type="checkbox"/>) Menos de 20% 2=(<input type="checkbox"/>) Mais de 20%
Q	Experiência de crédito no Banco	1=(<input type="checkbox"/>) Sim > 2 anos 2=(<input type="checkbox"/>) Sim < 2 anos 3=(<input type="checkbox"/>) Não
R	Histórico da conta corrente	1=(<input type="checkbox"/>) Normal 2=(<input type="checkbox"/>) Cheques devolvidos 3=(<input type="checkbox"/>) Cliente Novo 4=(<input type="checkbox"/>) Pequenos atrasos freqüentes
S	Sócios da empresa possuem restrições	1=(<input type="checkbox"/>) Sim 2=(<input type="checkbox"/>) Não
T	Sócios da empresa tiveram restrições baixadas nos últimos cinco anos	1=(<input type="checkbox"/>) Sim 2=(<input type="checkbox"/>) Não
U	Sociedade entre cônjuges	1=(<input type="checkbox"/>) Sim 2=(<input type="checkbox"/>) Não
V	Existência de bens imóveis em nome dos sócios	Valor Numérico
W	Existência de bens móveis em nome dos sócios	Valor Numérico
X	Risco atribuído pelo Banco	1=(<input type="checkbox"/>) A 2=(<input type="checkbox"/>) B 3=(<input type="checkbox"/>) C 4=(<input type="checkbox"/>) D 5=(<input type="checkbox"/>) E
Y	Resultado	1=(<input type="checkbox"/>) Adimplente 2=(<input type="checkbox"/>) Inadimplente

APÊNDICE B – ANÁLISE FATORIAL (AF) E ANÁLISE DE COMPONENTES PRINCIPAIS (ACP)

Análise Fatorial (AF) é uma técnica multivariada que objetiva explicar as correlações existentes entre um conjunto grande de variáveis em termos de um conjunto de poucas variáveis aleatórias não observáveis, denominadas fatores.

Quanto mais fortes forem as correlações entre algumas variáveis dentre o grupo inicial, mais nítida é a visualização do fator gerado. Variáveis agrupadas num mesmo fator possuem portanto alta correlação, enquanto que as variáveis de fatores distintos possuem baixa correlação.

Objetivos da Análise Fatorial

O objetivo geral é condensar a informação, contida num número de variáveis originais, em um conjunto menor de fatores com um mínimo de perda dessa informação.

Os objetivos específicos são:

- a) Observar um conjunto de dimensões latentes num grande conjunto de variáveis - Análise Fatorial do tipo R.
- b) Combinar ou condensar um grande número de observações em grupos - Análise Fatorial do tipo Q.
- c) Identificar variáveis apropriadas para uma posterior regressão, correlação ou Análise Discriminante.
- d) Criar um novo conjunto de novas variáveis em menor número, para substituir outro conjunto.

Duas variações são utilizadas na técnica e são chamadas de Análise Fatorial ou Análise de Componentes Principais. A primeira busca o conhecimento de dimensões subjacentes e a segunda busca a redução de variáveis às principais.

O Modelo Fatorial Ortogonal

Considerando-se \underline{X} o vetor de variáveis originais, com dimensão p , com vetor de média $\underline{\mu}$ e matriz de Covariância Σ , não necessariamente com distribuição normal, tem-se que o modelo fatorial de \underline{X} é linearmente dependente sobre algumas variáveis aleatórias, não

observadas F_1, F_2, \dots, F_m , (sendo $m < p$ necessariamente, caso contrário, não haveria ganho com a AF), que são denominados fatores comum, p fontes de variações aditivas, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ são os erros, o modelo de Análise Fatorial pode ser representado com notação matricial como:

$$X - \mu = LF + \varepsilon$$

onde:

\underline{X} é o vetor das variáveis originais para cada indivíduo;

$\underline{\mu}$ é o vetor das médias das i -ésimas variáveis;

L é a matriz dos pesos ou carregamentos nas i -ésimas variáveis e nos j -ésimos fatores F_j ;

\underline{F} é o vetor dos fatores comuns;

$\underline{\varepsilon}$ é o vetor dos erros ou fatores específicos.

Se o objetivo for o de identificar variáveis importantes para uso posterior, o analista deve examinar a matriz de dados e selecionar a variável com mais alto peso fatorial como representativa de uma dimensão particular. Se o objetivo for o de criar um conjunto inteiramente novo, com um número menor de variáveis, então os escores fatoriais devem ser calculados e utilizados como dados brutos em análises posteriores.

Comunalidades

São porções da variância de uma variável X , distribuída entre m fatores comuns.

Método das Componentes Principais para estimar os pesos fatoriais (carregamentos fatoriais)

Para investigar as relações entre um conjunto de p variáveis correlacionadas pode ser útil transformar o conjunto de variáveis originais em um novo conjunto de variáveis não-correlacionadas chamadas componentes principais, tendo propriedades especiais em termos de variâncias.

As novas variáveis: componente principais, são combinações lineares das variáveis originais e são derivadas em ordem decrescente de importância tal que, por exemplo, a 1ª componente principal é a combinação linear normalizada com variância máxima.

As componentes principais podem ser obtidas através dos autovalores da matriz de

correlação. Representa-se por $(\lambda_1^*, e_1^*), (\lambda_2^*, e_2^*), \dots, (\lambda_{20}^*, e_{20}^*)$, os pares de autovalores e autovetores da matriz de correlação. A variância total da população será igual a p e consequentemente a proporção da variância explicada pelo k -ésimo componente principal é dada por $\frac{\lambda_k}{20}, k = 1, 2, \dots, 20$.

Seja $m < p$ o número de fatores comuns. A interpretação dos fatores de uma AF é feita por meio dos pesos ou cargas fatoriais, que são parâmetros de um modelo de AF que expressam as covariâncias entre cada fator e as variáveis originais e é necessário escolher o número de fatores. Como escolher esse número de fatores:

- Experiência no assunto pesquisado (método subjetivo);
- Critério de Kaiser: tantos fatores quanto for o número de autovalores maiores que 1;
- Escolher o número de fatores que explicam uma proporção especificada da variância total.

Escores Fatoriais Estimados: em muitas aplicações precisamos estimar o valor de cada um dos fatores para uma observação individual $\underline{x} = [x_1 \ x_2 \ \dots \ x_p]'$. Estes valores dos fatores são chamados de escores fatoriais.

Os escores fatoriais consistem no produto matricial entre os valores observados para as variáveis e os pesos fatoriais.

Matriz de Resíduos

A Análise Fatorial também permite o cálculo dos valores residuais ou matriz dos resíduos. Essa matriz é outra forma de se avaliar se o modelo fatorial está próximo da realidade, pois expressa a diferença entre as correlações e o produto dos pesos estimados e sua transposta, acrescidos dos erros. Valores próximos de zero indicam que o modelo fatorial é adequado.

Rotação dos fatores

A rotação de fatores é uma técnica para girar os eixos de referência dos fatores, em torno da origem, até alcançar uma posição ideal. Ela pode ser ortogonal ou oblíqua, caso os eixos se mantiverem ou não em 90 graus entre si durante o giro. O objetivo é otimizar os

pesos fatoriais. Como o próprio nome já diz, a rotação consiste em girar os eixos em um ângulo θ , oferecendo uma nova estrutura para os pesos, de tal forma que cada variável tenha peso alto em um único fator e pesos mais baixos ou médios nos demais fatores. A rotação não produz uma estrutura visível quando $m > 2$, ou seja, o número de fatores é maior que dois e quando isso ocorre programas computacionais são usados para executar a rotação e proceder os cálculos dos fatores rotacionados, embora a visualização continue impossível.

O método de rotação mais utilizado foi desenvolvido por Kaiser e recebe o nome de Rotação Varimax ou Normal Varimax, o qual simplifica as colunas da matriz de fatores.

Uma confusão que deve ser esclarecida é que Análise de Componentes Principais (ACP) não é igual a Análise de Fatores. Embora os resultados de ambos os métodos sejam muitas vezes semelhantes, existem grandes diferenças conceituais entre os mesmos. Simplificadamente, a Análise de Fatores busca uma estrutura causal entre as variáveis, enquanto ACP busca apenas reduzir o conjunto de variáveis mantendo, ao máximo, o poder de explicação (variância total). Apesar disto, a Análise de Componentes Principais pode ser utilizada para se associar um componente principal a um fator real.

Interpretações

Soma dos quadrados das cargas fatoriais = autovalor = quantidade de variância explicada pelo fator.

Soma das comunalidades = total de variância extraída pela solução fatorial.

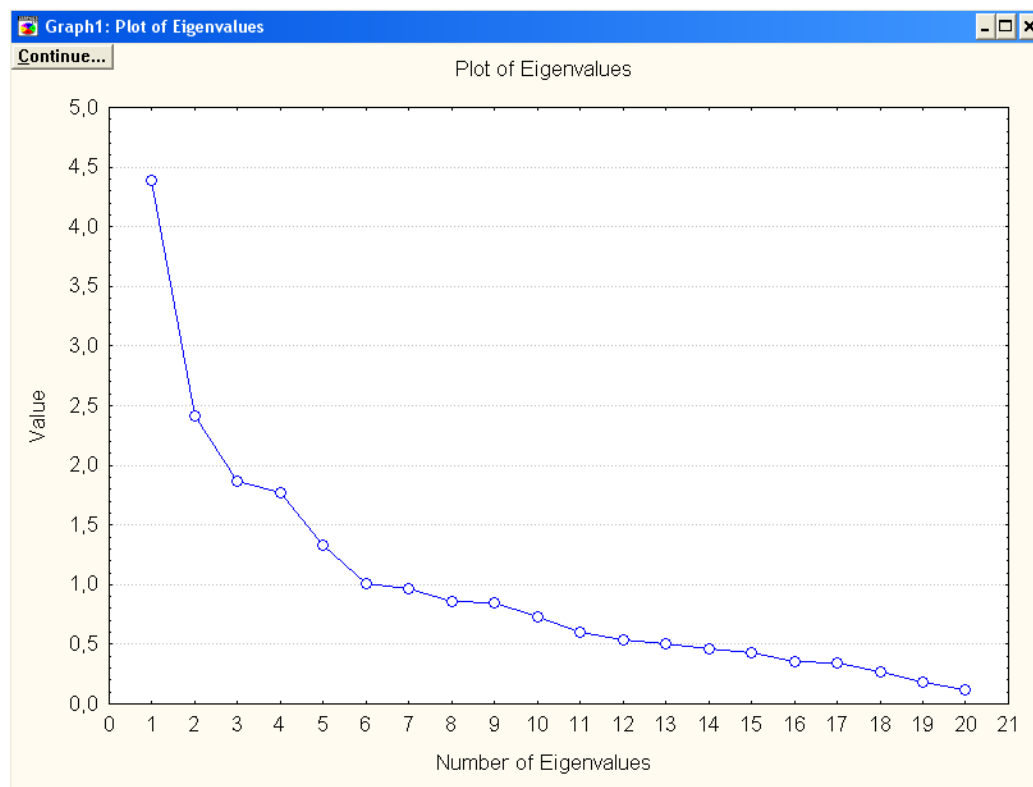
As comunalidades mostram o quanto de variância numa variável é explicada pelos fatores em conjunto.

Pode-se utilizar variáveis representantes para análises subseqüentes, com base nas suas cargas fatoriais elevadas - com rotação ortogonal.

Pode-se utilizar os escores fatoriais de cada observação como um conjunto inteiramente novo de variáveis independentes. Escores fatoriais elevados mostrarão que aquela observação tem alta influência daquele fator.

APÊNDICE C – AUTOVALORES E VARIÂNCIA EXPLICADA

Autovalores			
Extração de Componentes Principais			
	Autovalores	Variância % total	Variância Explicada
1	4,393195033	21,96597517	21,96597517
2	2,41403938	12,0701969	34,03617207
3	1,862138634	9,310693168	43,34686523
4	1,771249148	8,85624574	52,20311097
5	1,33282277	6,664113852	58,86722483
6	1,003875857	5,019379284	63,88660411
7	0,961028642	4,805143212	68,69174732
8	0,85529126	4,276456298	72,96820362
9	0,844341187	4,221705936	77,18990956
10	0,735118616	3,675593082	80,86550264
11	0,604247345	3,021236726	83,88673936
12	0,537756635	2,688783177	86,57552254
13	0,508394443	2,541972217	89,11749476
14	0,465542051	2,327710256	91,44520501
15	0,432117133	2,160585666	93,60579068
16	0,359784852	1,79892426	95,40471494
17	0,345771141	1,728855706	97,13357065
18	0,272293175	1,361465873	98,49503652
19	0,180364649	0,901823244	99,39685976
20	0,120628047	0,603140236	100

APÊNDICE D – GRÁFICO DOS 20 AUTOVALORES

APÊNDICE E – COMUNALIDADES

Comunalidades						
Extração dos Componentes Principais						
	Até o fator5	Até o fator 6	Até o fator 7	Até o fator 8	Até o fator 9	R
	Fator	Fator	Fator	Fator	Fator	Multiplo
X1	0,746323401	0,75658329	0,760786889	0,768482291	0,797830706	0,675203613
X2	0,681665048	0,693990442	0,723548392	0,723689616	0,730019445	0,648218368
X3	0,787171815	0,883620775	0,884068476	0,887382725	0,89278349	0,775093524
X4	0,725313626	0,741682505	0,75738762	0,760119104	0,76262909	0,332755439
X5	0,599102344	0,643898011	0,651177492	0,824403629	0,906697815	0,664036943
X6	0,663618148	0,727321881	0,730725116	0,765785585	0,794575346	0,488004006
X7	0,304304905	0,35619611	0,884348812	0,886266241	0,889814348	0,262288735
X8	0,464467126	0,512539174	0,514095782	0,531265755	0,676577727	0,261920181
X9	0,624480189	0,664293105	0,680925822	0,699640787	0,730315586	0,338710876
X10	0,491347362	0,668515024	0,784841021	0,820595218	0,837360042	0,387629208
X11	0,372797384	0,584169653	0,585016031	0,863206161	0,871746527	0,22305296
X12	0,4888791	0,659692388	0,67570176	0,848168453	0,937697251	0,51518802
X13	0,650288224	0,651421328	0,651875554	0,710683997	0,711410014	0,51674745
X14	0,610991523	0,613332711	0,617945805	0,617993992	0,618153057	0,529132451
X15	0,726555634	0,736693975	0,740273879	0,740627081	0,742533358	0,366780526
X16	0,710714061	0,742244065	0,743345251	0,744981223	0,746144134	0,405593749
X17	0,468152585	0,480488823	0,53087082	0,530894266	0,62637304	0,385756018
X18	0,647902811	0,648105092	0,727637737	0,727648474	0,751568182	0,580967177
X19	0,664687468	0,667548764	0,69938222	0,6993848	0,716286584	0,625797815
X20	0,344682215	0,344983705	0,394394985	0,442421326	0,697466168	0,182582444

APÊNDICE F – MATRIZ DE CORRELAÇÃO RESIDUAL

Correlação Residual										
Extração de Componentes Principais										
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	0,2022	-0,0390	0,0214	0,0219	-0,0418	0,0022	0,0426	-0,0031	0,0214	-0,0148
X2	-0,0390	0,2700	0,0048	-0,0105	0,0036	0,0052	-0,0423	0,0128	-0,0208	0,0305
X3	0,0214	0,0048	0,1072	0,0135	0,0268	-0,0383	0,0338	-0,0029	0,0408	0,0103
X4	0,0219	-0,0105	0,0135	0,2374	0,0591	0,0018	0,0618	-0,0732	-0,1388	-0,0044
X5	-0,0418	0,0036	0,0268	0,0591	0,0933	0,0034	-0,0173	-0,0142	-0,0736	0,0030
X6	0,0022	0,0052	-0,0383	0,0018	0,0034	0,2054	-0,0583	-0,0163	-0,0520	-0,1428
X7	0,0426	-0,0423	0,0338	0,0618	-0,0173	-0,0583	0,1102	-0,0806	0,0345	0,0027
X8	-0,0031	0,0128	-0,0029	-0,0732	-0,0142	-0,0163	-0,0806	0,3234	-0,1181	0,0922
X9	0,0214	-0,0208	0,0408	-0,1388	-0,0736	-0,0520	0,0345	-0,1181	0,2697	-0,0031
X10	-0,0148	0,0305	0,0103	-0,0044	0,0030	-0,1428	0,0027	0,0922	-0,0031	0,1626
X11	-0,0130	-0,0450	-0,0350	-0,0049	0,0538	-0,0106	0,0032	-0,0494	-0,0350	0,0018
X12	0,0138	-0,0488	-0,0524	0,0393	-0,0067	0,0092	-0,0035	0,0240	-0,0374	0,0173
X13	-0,0007	-0,0013	0,0069	0,0455	-0,0240	0,0157	-0,0268	-0,0326	0,0559	-0,0050
X14	0,0355	0,0540	0,0604	0,0106	-0,0326	0,0364	-0,0019	0,0162	0,0833	0,0552
X15	0,0164	-0,0059	0,0498	0,1177	0,0182	-0,0142	0,0123	-0,0274	-0,0136	-0,0380
X16	-0,0215	0,0232	0,0315	-0,0598	-0,0035	-0,0264	-0,0280	0,0312	0,0431	-0,0364
X17	0,0610	0,1182	0,0217	-0,0071	-0,0663	0,0111	0,0590	0,0312	-0,0623	-0,0604
X18	-0,0142	-0,1071	0,0107	0,0522	-0,0384	0,0335	0,1062	-0,0254	0,0159	-0,0559
X19	-0,1081	0,0044	0,0129	0,0018	-0,0102	-0,0256	-0,0269	-0,0645	0,0177	0,0131
X20	-0,0666	0,0355	-0,0373	0,0423	0,0329	0,0344	0,0220	-0,2563	0,0211	-0,0786
	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
X1	-0,0130	0,0138	-0,0007	0,0355	0,0164	-0,0215	0,0610	-0,0142	-0,1081	-0,0666
X2	-0,0450	-0,0488	-0,0013	0,0540	-0,0059	0,0232	0,1182	-0,1071	0,0044	0,0355
X3	-0,0350	-0,0524	0,0069	0,0604	0,0498	0,0315	0,0217	0,0107	0,0129	-0,0373
X4	-0,0049	0,0393	0,0455	0,0106	0,1177	-0,0598	-0,0071	0,0522	0,0018	0,0423
X5	0,0538	-0,0067	-0,0240	-0,0326	0,0182	-0,0035	-0,0663	-0,0384	-0,0102	0,0329
X6	-0,0106	0,0092	0,0157	0,0364	-0,0142	-0,0264	0,0111	0,0335	-0,0256	0,0344
X7	0,0032	-0,0035	-0,0268	-0,0019	0,0123	-0,0280	0,0590	0,1062	-0,0269	0,0220
X8	-0,0494	0,0240	-0,0326	0,0162	-0,0274	0,0312	0,0312	-0,0254	-0,0645	-0,2563
X9	-0,0350	-0,0374	0,0559	0,0833	-0,0136	0,0431	-0,0623	0,0159	0,0177	0,0211
X10	0,0018	0,0173	-0,0050	0,0552	-0,0380	-0,0364	-0,0604	-0,0559	0,0131	-0,0786
X11	0,1283	0,0105	-0,1286	-0,0658	-0,0635	-0,0049	-0,0603	-0,0342	-0,0404	0,0756
X12	0,0105	0,0623	0,0295	-0,0033	-0,0007	-0,0460	-0,0732	0,0276	-0,0505	-0,0382
X13	-0,1286	0,0295	0,2886	0,0004	0,0633	-0,0838	-0,0747	0,0060	0,0625	-0,0113
X14	-0,0658	-0,0033	0,0004	0,3818	0,0084	-0,0369	-0,1019	0,0123	0,0201	-0,0863
X15	-0,0635	-0,0007	0,0633	0,0084	0,2575	0,1620	-0,0189	-0,0020	0,0059	-0,0481
X16	-0,0049	-0,0460	-0,0838	-0,0369	0,1620	0,2539	0,0319	-0,0576	-0,0043	-0,0502
X17	-0,0603	-0,0732	-0,0747	-0,1019	-0,0189	0,0319	0,3736	0,0404	0,0208	0,0284
X18	-0,0342	0,0276	0,0060	0,0123	-0,0020	-0,0576	0,0404	0,2484	-0,0548	-0,0394
X19	-0,0404	-0,0505	0,0625	0,0201	0,0059	-0,0043	0,0208	-0,0548	0,2837	0,1424
X20	0,0756	-0,0382	-0,0113	-0,0863	-0,0481	-0,0502	0,0284	-0,0394	0,1424	0,3025

APÊNDICE G – ESCORES FATORIAIS

Escore fatorial									
Rotação: Varimax normalizada									
Extração de Componentes Principais									
	Fator	Fator	Fator	Fator	Fator	Fator	Fator	Fator	Fator
	1	2	3	4	5	6	7	8	9
1	0,664891	-0,974487	0,452728	1,381827	-0,461237	0,945732	-0,006558	13,416875	0,324411
2	0,221165	-0,145832	0,930734	0,660153	0,144635	10,444331	-0,391824	-0,623494	-0,065142
3	0,380235	-1,115832	1,088369	-0,629681	-0,808099	0,024864	-0,697410	0,082436	0,885414
4	-0,135935	-0,354389	1,010698	-0,595872	-1,133244	0,289969	-0,842243	-0,536147	0,715864
5	1,035935	-0,565457	0,732800	-0,754422	-1,034923	-0,430290	-0,320083	-0,164166	0,315541
6	0,363784	0,250881	1,026831	1,954804	-0,661587	-0,231886	-0,499138	-0,111928	0,381107
7	1,028159	-1,085857	0,837824	-0,725444	-0,877894	0,369741	-0,443357	0,184363	0,399711
8	1,131980	-0,762266	0,472222	0,351658	-1,312538	-0,644262	-0,256392	-0,089800	0,048997
9	-0,228996	0,849437	1,168714	-0,360235	0,855584	0,144127	0,748711	-0,177341	0,325197
10	-0,444810	0,453837	1,122313	-0,289188	-0,483492	-0,079632	-1,091833	-0,227640	0,667447
11	1,053074	-0,967695	0,706846	-0,748523	-0,898581	0,385176	-0,195845	0,086132	0,269862
12	0,783811	-0,911130	1,034083	-0,441449	1,453175	-0,450842	-0,392936	-0,186947	0,344317
13	0,500118	-1,131140	1,201488	0,426854	-0,772097	-0,172467	-0,670339	0,163544	1,066200
14	-0,521660	-0,133305	0,064826	-0,252904	-0,563029	6,730519	-1,827735	-1,170181	1,732093
15	1,188615	-1,168504	-1,109859	-0,759408	-0,731746	-0,276659	-0,199567	0,235612	1,356961
16	-0,217622	0,004871	-1,095415	-0,800958	0,661958	-0,108272	-1,060646	-0,059953	1,343819
17	0,401021	1,586126	-1,099737	-0,305437	0,351360	-0,431630	-0,936350	0,084649	1,317625
18	0,855544	-0,482177	0,944234	-0,658823	-0,835398	-0,515181	-0,518416	0,617959	0,566035
19	-0,664384	1,798840	1,000838	0,190641	-0,547444	0,206631	-1,249752	-0,215410	0,415981
20	-0,486136	2,480456	0,148113	0,138382	-1,064510	0,224682	-1,196984	0,340753	1,270763
21	0,321286	-0,077810	0,568900	0,303674	-1,142959	-0,130342	-0,532583	-0,131774	1,984945
22	0,677087	-1,442133	0,255757	0,614412	-0,954883	0,379253	-0,504654	-0,053254	0,701587
23	0,323994	0,147313	1,086746	0,762825	-0,555459	-0,225491	1,346535	-0,433138	2,046057
24	1,041662	0,195022	0,527703	-0,214772	-0,904811	0,360207	-0,000484	-0,060763	-1,461435
25	0,075614	-0,261560	0,344796	0,428093	-1,397915	0,292983	-0,386979	-0,778158	-1,477203
26	0,452132	0,149966	0,941937	-0,660785	-1,216417	-0,471739	-0,648478	-0,227532	0,719188
27	0,770710	-1,253501	-0,449462	2,112122	-0,876346	-0,512564	-0,633083	0,019099	0,227878
28	0,188634	0,088408	0,632775	1,832607	-0,702459	-0,332359	-0,419115	-0,408234	1,816538
29	-0,002992	0,070777	0,723296	-0,511556	1,148123	-0,328059	1,045263	-0,337311	1,505437
30	-0,174745	-0,499243	-0,289649	2,407942	0,769459	-0,151144	-1,118835	-0,656204	0,327649
31	1,078942	0,243024	0,617272	-0,752114	-1,235727	0,390922	-0,178848	-0,199815	0,105960
32	-0,107081	-0,530270	1,012936	-0,608213	-1,125180	-0,070376	-0,881043	-0,376171	0,782105
33	0,430842	1,070852	0,289567	-0,854027	-1,092727	-0,529428	-0,530077	0,007521	1,568693
34	-0,029331	1,176105	0,793128	-0,361692	-0,638093	-0,178522	1,110088	-0,240913	1,628291
35	-0,130901	1,133337	1,102169	-0,236730	-0,906127	-0,003536	-0,741054	-0,470448	0,674046
36	1,216375	1,455913	0,275453	1,628464	0,739347	-0,603391	0,036368	-0,098532	1,082314
37	0,885014	-0,549560	0,580653	-0,731900	-0,663790	-0,425362	-0,068614	-0,063772	-1,353796

38	0,016828	0,635056	0,458896	1,124863	-0,140072	-0,256034	0,994044	0,192983	1,157170
39	1,097750	1,154029	0,942083	-0,632118	-1,415998	-0,263256	1,528533	-0,097099	0,155154
40	0,640311	2,633294	0,278931	0,395211	1,195516	-0,147443	-0,517993	0,455756	0,973089
41	1,100935	-0,446186	0,409688	1,437939	-0,981505	-0,038084	0,062480	-0,247282	-1,586213
42	0,584967	-1,127748	-0,414755	2,490463	0,083677	-0,253584	-0,268505	-0,170391	-1,433853
43	0,943014	-0,819234	0,808708	-0,712402	-0,843610	-0,077880	-0,300344	0,075121	0,374117
44	0,867066	-0,976999	-1,783989	0,616393	-0,720729	0,527719	-0,751490	0,281271	0,293686
45	0,370584	-0,875931	-0,630582	2,447402	-0,514284	1,440611	-0,099328	-0,682072	-0,246616
46	0,786499	-1,241975	0,471403	1,615865	1,231928	-0,361780	-0,096502	-0,389461	1,605232
47	1,169587	-0,363243	0,499898	-0,914280	0,365825	-0,186744	-0,297480	0,021525	0,005312
48	-0,164630	-0,167370	0,473234	0,340823	2,014603	-0,073053	-0,943649	-0,017128	1,391476
49	1,148751	-0,712738	0,232149	1,289154	0,608619	0,071939	0,127494	-0,334767	-1,861206
50	0,759903	-1,200702	0,525061	-0,621353	1,285122	-0,647966	-0,267475	-0,465949	1,721304
51	0,973002	-1,117338	0,508025	-0,750660	-0,553669	0,412574	0,078397	0,276032	-1,412234
52	0,791966	-1,242185	0,492827	1,614714	1,212387	-0,508691	-0,160330	-0,382945	1,639163
53	0,995712	-1,356219	-0,744341	-0,352103	0,510303	-0,617335	-0,675246	-0,385628	-0,086449
54	-0,386777	-0,555343	0,807937	1,624443	1,768224	0,021330	-1,271127	0,357539	0,136418
55	0,947013	-0,863769	0,705923	-0,848468	0,693696	-0,489618	-0,393074	-0,126887	0,180964
56	0,941630	-0,970603	-0,970179	-1,197282	1,039474	0,026725	-0,271141	0,535592	-1,570497
57	1,056570	-0,948860	0,708676	-0,750316	-0,911821	0,322214	-0,210748	0,073342	0,275413
58	0,813934	-0,460715	1,057095	-0,332252	-0,241607	-0,422102	-0,242993	-0,302509	0,451686
59	0,703880	-0,758871	0,817942	-0,448006	1,704704	-0,478137	-0,124890	-0,242949	-1,390691
60	0,642489	1,984585	-0,775560	-0,758460	0,766059	-0,316180	1,350610	0,106846	1,149078
61	0,828625	0,041019	0,671708	1,765517	1,333458	-0,250483	0,008854	-0,191338	-1,607327
62	0,658268	-0,339562	-0,821015	-1,101075	1,172527	0,066797	-0,265843	0,197474	-1,560637
63	1,012587	-0,447132	0,457958	-0,780786	-0,823410	-0,238079	0,009432	-0,247302	-1,482773
64	0,903610	0,778817	0,736688	1,228774	0,788988	-0,405522	-0,253101	-0,179719	-0,052148
65	-0,189660	-1,011982	-0,275607	2,393465	0,889145	-0,323618	-1,086232	-0,937168	0,343486
66	0,823170	0,836228	0,649530	-0,467272	1,204513	-0,343437	-0,095755	-0,186232	-1,596891
67	1,333243	-0,352727	0,022495	-1,029117	0,569393	-0,250462	0,097205	-0,049399	1,083827
68	1,232201	-0,616162	-1,308522	-1,327376	0,798854	0,217433	-0,102306	0,057363	1,085884
69	0,886489	1,084649	0,316244	0,433993	-1,284795	-0,538311	-0,059644	0,011743	-1,732991
70	1,174141	-0,851684	-1,436840	-0,894149	0,841135	0,998729	0,269321	-0,381688	0,745178
71	0,996328	-0,033260	0,538779	0,375336	0,500385	-0,694936	1,691606	-0,051232	-1,848072
72	0,084992	-0,432526	-1,512367	2,095888	0,872659	-0,390189	-0,897851	-0,524278	1,408260
73	0,839275	0,267623	-1,995249	0,545370	-0,535768	0,045873	-0,306861	0,003513	-0,630418
74	1,270367	-0,201613	-1,362993	-0,915000	0,493020	-0,329808	-0,172421	-0,149242	0,944701
75	0,998368	0,718117	0,229936	-0,582313	1,403893	-0,711569	0,153937	-0,138231	-0,417974
76	1,047934	-0,094491	-2,867629	-0,771072	0,807174	-0,363479	-0,243166	-0,101078	-0,831420
77	0,442721	2,940324	-1,754204	0,396117	1,282309	0,069845	-0,824339	0,593362	0,838621
78	0,879404	-0,070276	0,453897	-0,842167	0,797244	-0,283408	-0,119337	-0,010233	-1,613701
79	1,042198	0,521420	0,735154	-0,513169	0,866087	-0,175559	-0,248962	-0,093965	0,046238
80	-0,744263	0,932186	-0,855968	2,146178	0,886628	0,458768	-1,668280	-0,160464	-0,441983
81	0,848667	-0,183596	-0,995390	-1,046179	-0,551161	1,027029	0,171253	0,147154	-1,662344
82	1,037222	-0,147973	0,209489	-0,610172	1,584181	-0,587033	0,245877	-0,186978	-0,406584

83	1,317076	-0,178864	0,634714	1,688493	1,096897	-0,707665	1,851396	-0,192435	1,220229
84	0,618841	0,270719	-1,694144	-0,844808	-1,091633	-0,232151	-0,549587	-0,139550	1,591417
85	0,800597	0,752868	-0,388605	2,463568	-0,833165	-0,415781	-0,559492	-0,403780	0,047489
86	-0,206750	0,470545	0,733903	-0,725220	-1,314969	-0,158646	-1,066119	-0,186325	0,466853
87	1,265754	-0,612258	0,635454	1,497726	-1,021473	-0,114522	1,925720	-0,130404	-1,599987
88	0,527629	1,231634	0,817142	1,900761	-1,065988	-0,078417	-0,379532	-0,306460	0,143219
89	0,583978	0,787562	0,944614	0,144439	-0,962272	-0,284856	2,264399	0,058688	-0,131395
90	-0,486425	0,187784	0,785001	-0,434170	1,398824	0,014856	-0,851805	-0,246684	-1,273904
91	0,954992	0,429525	-1,318045	-1,250771	0,990781	-0,761540	-0,147363	0,199953	-0,433106
92	1,182286	0,542012	0,471690	-0,903754	0,124920	-0,077363	-0,353398	-0,017534	-0,073103
93	0,068328	0,847082	1,181431	0,760345	0,957168	0,144730	1,267468	-0,437729	-1,093176
94	0,601611	-0,554769	-1,626585	-0,851045	-0,888319	-0,429912	-0,592690	-0,162160	1,700984
95	0,205536	0,462752	0,541704	-0,763774	-0,873276	-0,149413	-0,567150	-0,131177	-1,264328
96	0,661032	1,148574	0,898474	-0,615207	-1,618635	-0,141929	1,287187	-0,124459	0,146915
97	0,027559	0,886544	-0,034284	-0,089903	-0,780071	-0,329625	-0,600606	-0,630206	1,695234
98	0,843014	0,963608	0,842861	0,631063	1,309007	-0,240983	-0,107493	-0,189493	-1,406166
99	0,102678	-0,889332	1,492213	0,756316	1,128262	0,041810	1,024912	-0,493338	0,788468
100	0,187878	0,467260	0,212566	-0,743509	1,962780	-0,357718	-0,774912	0,446580	0,884288
101	1,440671	-0,903041	-1,167839	-0,759432	0,862635	2,236339	4,530315	-0,438726	0,372880
102	1,016609	1,455142	1,083554	-0,369681	0,743821	-0,272755	1,491730	-0,033603	0,052356
103	0,124254	-1,120521	0,895173	-0,477730	0,937848	0,213963	-0,688533	-0,530538	0,493970
104	0,963582	0,029586	0,249231	-0,584756	1,614480	-0,551510	0,244026	-0,288864	-0,407671
105	-0,181000	-0,240600	0,779376	-0,620423	-0,921922	-0,177953	-0,605214	-0,541763	-0,979123
106	0,473852	-0,762986	-1,863184	-0,381770	-0,379363	-0,671599	-0,849407	0,347123	-1,290216
107	0,734093	-0,584563	0,704982	-0,672011	-0,487188	-0,305281	-0,111728	0,051212	-1,266251
108	1,042143	-0,881896	0,323241	-0,915685	0,834475	0,058974	0,035694	-0,204349	-1,716983
109	0,179751	0,368609	0,603018	0,366057	-1,228855	0,721458	-0,304467	-0,493802	-1,142931
110	1,091902	-0,520362	-1,932013	-0,897019	-0,698192	0,688067	0,063189	-0,006144	1,075247
111	1,032791	-1,725252	-0,904265	1,887713	0,618028	0,809228	-0,181936	-0,134205	-0,361949
112	-0,737915	0,263162	-0,615892	-0,831045	2,060364	0,775386	-1,412727	0,651115	-0,219653
113	0,889358	-1,368231	-0,934806	-0,342013	0,850520	-0,286746	-0,387411	-0,321783	-1,814095
114	1,246196	-0,991230	0,183923	-1,007997	0,767572	-0,643418	-0,089353	-0,171613	1,261517
115	1,116795	-0,334570	1,295829	0,677872	1,238048	-0,083617	1,541845	0,024761	0,360974
116	1,204901	-1,209755	0,639301	1,439821	-1,075784	-0,181949	-0,192097	0,318601	0,267879
117	-0,050718	0,176301	-2,088949	0,779493	-1,188268	-0,203631	-0,430762	-0,580449	-0,519126
118	0,460205	2,828551	0,797485	0,601396	1,271512	0,258507	1,289009	0,433574	-1,912417
119	0,928275	0,596868	-1,143118	0,933077	-1,453485	-0,539080	-0,305675	-0,256286	-2,022407
120	1,065177	-0,714149	-1,299649	-0,190958	-0,774421	2,883317	4,368674	-0,475779	-0,736390
121	-0,927112	2,172064	0,689962	-0,534445	1,086596	0,318416	-1,563251	0,241475	-0,048554
122	0,857748	1,649949	0,051773	-0,598297	0,858669	0,294311	-0,234431	0,091775	-0,055588
123	0,793251	0,455448	-1,301569	-0,722581	-1,056249	-0,124243	-0,627972	0,138138	0,150683
124	-0,096272	3,170874	-0,650546	0,955127	0,667051	0,291403	0,644501	0,686024	-0,531629
125	1,069633	-0,088627	-1,580910	0,811158	-1,068053	-0,684770	0,104192	-0,461292	-0,889331
126	-2,045335	0,404431	0,057967	-1,192438	0,616095	0,007878	-0,833908	0,357426	-0,901274
127	1,043769	0,172896	0,412217	-0,777776	-0,992230	-0,237559	0,002410	-0,047300	-1,522689

128	1,025686	0,034243	-2,968640	-0,768702	0,850325	0,232769	0,029919	-0,232697	-0,999728
129	0,964038	0,875946	-1,714552	-0,904009	0,310989	0,915380	-0,246795	0,064338	-0,424342
130	1,273520	1,362829	-1,328633	-1,198395	-1,283482	-0,108293	-0,054754	0,165426	1,145358
131	1,021143	-0,515301	0,482444	-0,782550	-0,826933	-0,415013	-0,057601	-0,183628	-1,433099
132	0,345441	2,724684	-2,014178	-0,448194	0,151582	-0,176906	-0,952133	0,477024	0,840504
133	1,080625	2,101631	-1,758245	-0,303915	-1,336515	-0,626957	-0,436759	0,191046	1,264588
134	0,855244	1,174549	0,427576	0,998286	-1,186406	-0,367746	-0,034528	0,006251	-1,644332
135	0,311120	1,621137	-0,459952	-0,593084	-0,800775	0,018542	-0,679100	0,025408	0,260776
136	-1,401713	0,019453	0,395248	-0,425475	-1,327813	-0,058523	-0,282112	0,182003	-0,281431
137	-0,461595	-0,255325	0,598935	0,455708	1,044848	-0,198714	1,732321	-0,002360	1,227470
138	-1,790997	0,372268	0,177067	-0,744047	-0,154314	-0,040629	1,478893	0,411228	0,494024
139	-1,727503	-1,127226	-0,762916	1,680074	-0,612636	-0,369863	-0,508569	-0,252260	-0,510512
140	-1,282955	-1,077737	-0,938183	2,136886	-0,590277	-0,720559	1,594587	-0,381222	0,647316
141	-1,488146	0,461907	0,512569	1,505745	1,381965	0,257749	-1,335321	0,153536	-0,436770
142	-1,002832	0,343524	0,074108	-0,827690	0,683527	-0,219258	1,924743	-0,021794	0,461126
143	-1,422377	0,202825	-1,153131	-0,166998	0,079910	0,226247	1,505623	0,324296	-0,952185
144	-1,468629	0,262315	-2,373867	0,490214	-1,751168	-0,435614	1,053152	-0,017821	-0,901775
145	-1,653270	1,232053	0,358758	-0,698275	1,643313	0,245184	-1,213337	0,289554	-2,080408
146	-1,516060	-0,731442	0,621176	1,641204	-0,613565	-0,295828	-0,225400	-0,018507	-0,233488
147	-1,423961	0,781047	0,799998	-0,493318	-0,897978	-0,042412	1,609216	0,114516	-0,347739
148	-0,594724	0,550752	0,478727	-0,665071	1,035796	-0,257058	1,169586	-0,111355	1,035595
149	-1,497554	0,893777	0,317680	-0,930367	-1,511759	0,004723	-0,292890	0,185776	-0,427968
150	-1,479598	0,528901	-1,543768	-0,685499	1,069435	0,927919	0,258955	-0,042900	0,211092
151	-1,220532	2,264411	0,573143	0,621543	-0,616393	0,194908	-0,377617	0,222509	-1,693380
152	-1,515474	-0,274393	0,158825	0,181104	-1,476178	-0,431224	-0,206492	-0,305816	-0,626563
153	-1,945667	1,169248	-2,115644	-0,875066	1,387586	0,874120	1,374022	0,523984	-0,068982
154	-1,100590	-0,325711	-0,063297	-0,343573	-0,530643	-0,421483	-0,184183	0,023120	-0,329361
155	-1,339077	-0,052749	-1,222980	-0,105770	1,302238	-0,348476	1,834857	-0,144140	-1,124100
156	-1,051879	0,644344	0,309677	0,293929	0,810304	-0,156225	1,895621	-0,120346	0,661518
157	-1,173991	-0,741587	0,807693	-0,447697	0,156341	-0,070844	-0,735204	0,026980	-1,321839
158	-1,579571	-1,153257	0,572310	-0,707491	1,188103	-0,255283	-0,335989	0,015543	-0,328400
159	-0,947769	-0,197233	0,717212	-0,617684	1,152666	0,075844	-0,906602	-0,132695	-0,018414
160	-1,039435	-0,968972	0,851732	-0,594865	1,416317	-0,117341	-0,996493	-0,154000	0,134528
161	-0,468886	1,728003	0,694762	0,470587	0,972736	0,211095	-0,669011	0,127732	-0,349808
162	-0,593981	-0,882605	-2,462727	-0,490686	1,737183	-0,370238	-0,917814	0,086548	0,663067
163	-0,719089	1,324048	0,192509	-0,727381	0,902175	-0,262383	-0,682489	-0,014220	1,007693
164	-1,090454	0,783207	0,062076	-1,012984	-1,478746	-0,436083	1,833588	0,162109	0,664079
165	-0,926914	-0,533542	0,649606	-0,856667	-0,944708	-0,104297	-0,949339	-0,113431	0,167433
166	-1,419995	-1,732775	0,349591	-1,098631	0,595364	-0,353527	-0,485568	0,593822	-0,304484
167	-0,922365	-1,044395	0,726155	-0,848127	-0,817367	-0,231110	-1,064222	0,312187	0,320504
168	-1,327924	-1,016282	-0,128806	-1,070374	-0,812963	0,226741	0,259912	0,216362	0,752747
169	-1,463437	-0,720377	0,374485	-0,970745	-1,180355	-0,307099	-0,306601	0,060991	-0,284522
170	-0,611825	-1,025728	-0,750926	-0,317764	-0,782130	-0,349355	-1,093040	0,047683	-0,307417
171	-1,042554	0,482740	-0,652060	2,189229	0,754676	-0,158488	-1,231049	-0,449193	-0,348570
172	-2,010424	-0,598802	0,612364	-0,749996	1,489986	0,081095	1,070069	0,303437	-0,732268

173	-1,276561	-1,661756	-1,592100	-1,113134	0,735192	0,007888	-0,211373	0,430281	0,532292
174	-1,464491	-0,594009	0,350200	-0,973281	-1,209442	-0,259730	-0,265293	-0,044993	-0,328633
175	-1,385864	-0,800969	-1,499996	-0,958323	-0,926205	0,007765	-0,065312	0,069755	0,631327
176	-1,080321	-0,528329	0,964295	-0,456020	-0,213440	-0,182649	-0,945577	-0,211753	0,324260
177	-0,619199	-0,690510	0,656705	-0,556506	-0,213262	-0,382116	1,234262	-0,248931	1,358928
178	-0,627522	-0,814392	0,751040	-0,612659	1,567341	-0,236463	-0,758816	0,047649	-0,212676
179	-1,475777	-1,173698	0,499538	1,502015	1,022079	-0,192365	-0,252648	0,038914	-0,434763
180	-1,703411	0,471939	-0,722336	-0,852115	-0,568944	-0,172563	-0,508168	0,352839	-0,204447
181	-1,049801	-0,011876	0,070938	-1,040332	-1,339299	-0,456105	1,838920	0,061260	0,705423
182	-0,818887	-1,265192	0,543598	-1,013179	0,682134	-0,087644	-1,079374	0,140184	0,035983
183	-0,733694	1,490666	1,021294	0,556101	0,784387	0,479195	0,980775	-0,035151	-0,052665
184	-0,926571	1,090555	0,771948	1,759527	-0,816358	0,007057	-0,845034	-0,122061	0,042450
185	-1,186324	-1,207091	0,313309	-0,784068	1,227646	-0,505804	1,819195	-0,062899	0,732072
186	-1,103916	-1,374199	0,480329	0,296515	1,353891	-0,280181	1,818759	0,125119	0,935325
187	-1,404464	-0,392810	-2,374103	1,847418	-0,541122	-0,272813	1,556531	-0,298218	0,508602
188	-1,147056	-1,121470	0,620355	0,440398	-0,221901	-0,333297	1,837465	0,221412	1,174464
189	-1,092155	-0,458074	0,934978	-0,453004	-0,202657	0,056179	-0,860125	-0,279801	0,263551
190	-1,402242	-0,643254	-2,315683	1,853127	-0,488582	-0,401353	1,449287	-0,086657	0,608458
191	-1,389345	-0,705073	1,063411	0,575188	-0,435124	-0,015354	1,592439	-0,024116	-0,027774
192	-1,786497	-0,933228	-0,614092	0,005902	-0,472486	-0,439850	-0,622574	-0,387887	-0,333546
193	-1,650317	-1,208231	-1,155529	-0,416674	-0,895338	-0,272078	-0,365239	-0,049099	-2,180225
194	-1,617091	0,385399	0,125825	-0,936902	-1,053035	0,181131	0,058882	-0,045636	-2,157202
195	-1,462020	-0,032768	0,830379	-0,492050	-0,644925	0,174288	1,699207	0,064492	-0,315816
196	-1,552539	-0,428703	-2,552384	1,784851	-0,484829	-0,516791	-0,425911	-0,235232	0,626468
197	-1,602115	0,108880	0,605518	-0,563777	-0,639961	-0,099897	-0,214940	-0,017336	-0,260101
198	-1,159121	0,891784	0,067526	0,492735	-1,083242	-0,517772	1,995882	-0,117759	0,446700
199	-0,560889	0,849311	0,888930	0,597930	-0,307544	0,042324	-0,591257	-0,260537	-0,066044